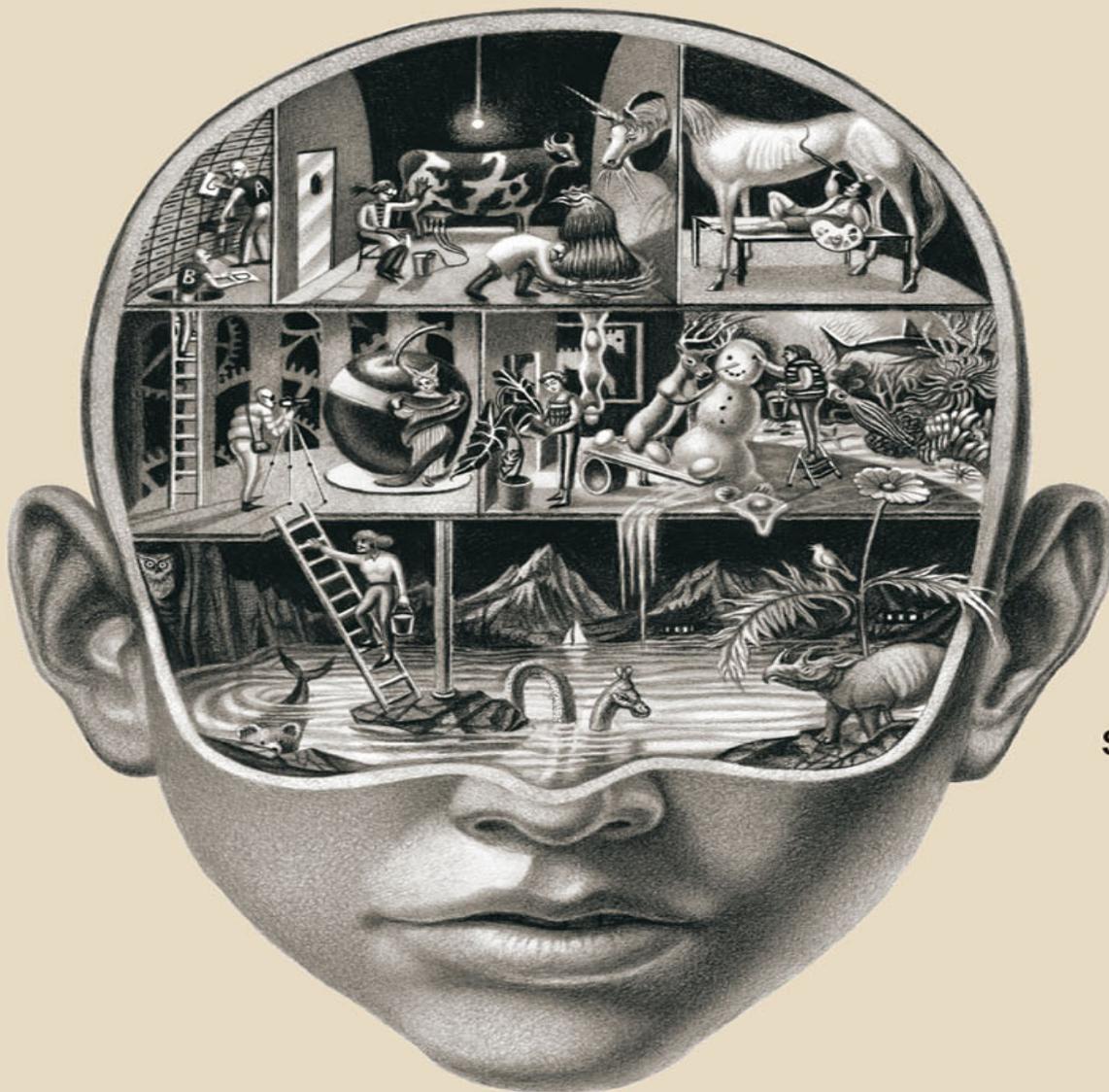


Roger Bartra

Chamanes y robots



Reflexiones
sobre el efecto
placebo y la
conciencia
artificial


ANAGRAMA
ARGUMENTOS

Índice

PORTADA

PRÓLOGO

I. LOS RITUALES DEL PLACER Y LA PALABRA

1. EL PLACEBO
2. LAS LIGADURAS DE QUSTA IBN LUQA
3. LOS PODERES MÁGICOS
4. UN VIAJE CHAMÁNICO EN BUSCA DEL ALMA PERDIDA
5. NEUROLOGÍA DEL EFECTO PLACEBO
6. SOBRE AMULETOS ELECTRÓNICOS Y CATARSIS
7. ZOMBIS Y TRANSHUMANISTAS

II. LA CONSTRUCCIÓN DE UNA CONCIENCIA ARTIFICIAL

1. EL MISTERIO DE LAS MÁQUINAS PENSANTES
2. EL EFECTO ROBÓTICO
3. ¿CÓMO SE EDUCA UN ROBOT?
4. PANPSIQUISMO
5. UNA CONCIENCIA MECÁNICA
6. LA CULTURA ROBÓTICA
7. PRÓTESIS Y SÍMBOLOS
8. EXPERIENCIAS ROBÓTICAS
9. LA EMANCIPACIÓN DE LOS EXOCEREBROS
10. MÁQUINAS SENTIMENTALES
11. LA PRUEBA DEL PLACEBO

NOTAS

CRÉDITOS

PRÓLOGO

La gran expansión de la inteligencia artificial eliminará miles de empleos en el futuro cercano y está sustituyendo desde hace tiempo a trabajadores por robots que realizan con mayor eficacia y rapidez tareas que hacían los humanos. Estamos viviendo una profunda revolución tecnológica. Se está generalizando el temor de que estas máquinas inteligentes adquieran formas superiores de conciencia y que incluso logren absorber a mentes humanas que funcionarían con soportes artificiales y sin cuerpos biológicos. Hay quienes tienen esperanzas de que este futuro robótico genere nuevas formas de vida humana liberadas de su estrecha condición biológica. En realidad, los humanos desde su origen se han apoyado en prótesis artificiales que han expandido extraordinariamente los límites biológicos de la conciencia. Pero hoy hay un extraordinario desarrollo de sofisticadas máquinas dotadas de inteligencia artificial.

En este ensayo quiero explorar la extensión de las funciones de la conciencia en las redes culturales que los humanos han tejido y lo quiero hacer partiendo del ejemplo de la manera en que los rituales influyen en las texturas cerebrales. Quiero explorar los ritos chamánicos, desde sus expresiones antiguas o tradicionales hasta las formas en que se presentan en la medicina moderna. En el núcleo de estos rituales se encuentra un fenómeno, el llamado efecto placebo, que cada día es más reconocido y estudiado. El efecto placebo implica una estrecha relación funcional entre los artificios de la cultura y las operaciones biológicas del sistema nervioso central. Por ello, podemos encontrar en este fenómeno claves importantes sobre el funcionamiento de la mente humana conectada a la artificialidad de rituales y prótesis de índole cultural. Estas son precisamente las claves que los ingenieros quieren descifrar para intentar que los robots funcionen con estructuras conscientes similares a las humanas. Así que hay cierto paralelismo entre los procesos que desencadena un chamán o un médico en la mente de los enfermos que quieren sanar y los mecanismos que construye un ingeniero para dotar a un robot de algo semejante a la conciencia. En ambos casos se trata de una ligazón entre procesos artificiales y conductas

conscientes. Si los robots van a ser conscientes, deberán ser contruidos como sistemas mixtos, a semejanza del carácter híbrido de la conciencia de que están dotados los humanos.

Hay dos temas que permiten relacionar el fenómeno chamánico con la robótica: el sufrimiento y el placer. Los rituales de sanación de médicos y chamanes están encaminados a disminuir o eliminar el dolor y los tormentos que padecen los humanos, y a proporcionarles placer y bienestar. Por su parte, la inteligencia artificial y los robots son contruidos, en principio, con el objetivo de reducir las fatigas y las penas que afligen a los humanos cuando trabajan. El trabajo, como la enfermedad, es una fuente de dolor y penuria. No quiero decir que el trabajo sea una enfermedad, pero sin duda ambos son condiciones que generan sufrimiento. El problema al que se enfrentan los constructores de robots es que sus máquinas carecen de sensibilidad, y sin ella parece difícil que pueda existir la conciencia. Estamos ante una paradoja: la conciencia está sustentada en el sufrimiento, pero los humanos estamos empeñados en aliviarlo o incluso eliminarlo. Los robots que hoy conocemos son máquinas insensibles que no sufren, y por ello mismo no parece que puedan tener conciencia.

Hay que destacar la otra dimensión de las emociones y los sentimientos, el placer, que se conecta con la exploración de la conciencia. Muchas de las prótesis que extienden la conciencia hacia las esferas sociales están concebidas para complacer, aliviar y dar placer. Y el efecto placebo que trato en la primera parte de este libro se refiere directamente a la manera de proporcionar placer, además de alivio, a los seres humanos. Desde fines del siglo XVIII la literatura médica comienza a reconocer el placebo como un medicamento administrado al enfermo más para darle placer que para que resulte beneficiado. Más adelante se comenzará a ver que el placer mismo puede ser un beneficio, aunque la sustancia recetada al paciente sea inocua. Pero los robots que conocemos y sobre los que reflexiono en la segunda parte de este libro son totalmente ajenos al placer y a las emociones. ¿Qué puede significar esta condición insensible para quienes se dedican a construir inteligencias artificiales y pretenden dotar a las máquinas de una conciencia artificial? En las páginas que siguen inicio una exploración sobre algunas dimensiones artificiales de la conciencia que no abordé en mi libro *Antropología del cerebro: el efecto placebo y la conciencia de los robots*.¹ Me parece que son aspectos muy significativos sobre los cuales hay

que reflexionar y que pueden contribuir a entender el misterio de la conciencia humana.

Hay quienes están convencidos de que la conciencia es una mera ilusión provocada por el funcionamiento de nuestro cerebro, y que por lo tanto no hay ningún misterio que investigar. La conciencia sería una especie de truco generado por la actividad neuronal. En contraste, desde la perspectiva del pansiquismo, hay quienes creen que hay conciencia en todo el universo y que su expresión en los humanos es solamente una variante compleja de un fenómeno cósmico. Otros postulan que la conciencia es una fuerza fundamental, similar a la gravedad. Hay teorías que buscan anclarse más en la realidad y plantean que la conciencia es un fenómeno cuántico o un proceso físico cuya naturaleza todavía no conocemos. En la misma línea se cree que la conciencia debe ser un fenómeno físico emergente que provoca un salto a una forma de gran complejidad.

La explicación de la conciencia más extendida es, desde luego, la religiosa, que la considera una entidad espiritual y metafísica.

Yo parto de la idea de que la conciencia es un fenómeno híbrido singular propio de los humanos. Con más precisión, diría que se trata de la autoconciencia. Es la unión de dos esferas, la cerebral y la cultural, cada una de las cuales responde a leyes diferentes y cuya confluencia es todavía un fenómeno no bien explicado. Mi propuesta se enfrenta a un problema: confluyen aquí dos expresiones científicas que responden a muy diferentes realidades, la neurología y la antropología. Son dos esferas científicas, las ciencias naturales y las ciencias sociales, que han alimentado desde hace mucho tiempo una desconfianza mutua. Las intromisiones de la biología en la sociología no han dado resultados muy fructíferos y han provocado cierto rechazo de quienes estudian los fenómenos sociales y culturales. Por otro lado, quienes parten de disciplinas duras, como la física, la química y la biología, suelen dudar del carácter científico de la antropología y la sociología. El abismo que separa a las ciencias naturales de las sociales es más grande que el que divide, por ejemplo, a la física cuántica de la gravitacional. Somos todavía víctimas de esa antigua segregación que, como afirmó C. P. Snow, separó la cultura científica de la humanista.

El lector de este libro podrá comprobar que los temas abordados

requieren unir la dimensión biológica con la cultural. Cuando se agrega el problema robótico nos topamos además con las dificultades de vincular la mecánica cibernética con las dimensiones biológicas y sociales. Hoy en día pocos reconocen la necesidad de trabajar en la unificación y articulación de dominios científicos tan alejados. Estas dificultades le dan a este libro el carácter ensayístico de una búsqueda en territorios llenos de paradojas.

I. Los rituales del placer y la palabra

Antropología del efecto placebo

1. EL PLACEBO

Quiero invitar al lector a un viaje que se aventura por territorios escabrosos pero intelectualmente estimulantes y, espero, divertidos. Es un camino riesgoso y retorcido que pasa por las ideas de un médico árabe del siglo IX, por los lamentos artificiales de los sicofantes medievales que se alquilaban para llorar a los difuntos y que nos conduce hacia los chamanes indígenas del Chaco paraguayo. Después las peripecias continúan hasta llegar a los curanderos kunas de Panamá estudiados por los antropólogos y los rituales psicoanalíticos encaminados a curar la esquizofrenia, para después comentar las investigaciones neurológicas sobre el efecto placebo. Llegaremos también a una breve discusión sobre la catarsis, los amuletos electrónicos modernos y el transhumanismo.

¿Qué me propongo en este viaje? Busco pruebas que demuestren que para entender el fenómeno de la conciencia humana es necesario escapar del cráneo que encierra al cerebro. Es una expedición antropológica que explora un tema inquietante: el llamado efecto placebo. Me interesé en el tema de la conciencia hace mucho y la búsqueda me llevó a escribir un libro, *Antropología del cerebro*, que se publicó en 2006. Pero una vez publicado me di cuenta de que no había examinado un fenómeno —el placebo— que es una prueba que apoya mis interpretaciones sobre los vínculos del cerebro con su entorno social y cultural. En el curso de las investigaciones me he topado con una gran resistencia de algunos neurocientíficos que no admiten que la conciencia pueda estar parcialmente ubicada fuera del cerebro, en las redes simbólicas de la sociedad. La conciencia, creen, no puede ser un fenómeno externo al cerebro y no admiten que pueda ejercer influencia sobre la actividad neuronal.

La idea de que la conciencia de los humanos es incapaz de influir en el funcionamiento del cerebro proviene de la afirmación de un principio científico fundamental: el mundo físico está cerrado causalmente y en consecuencia ninguna entidad espiritual o metafísica puede interferir en las operaciones del sistema nervioso central. Muchos neurocientíficos ven con sospecha el estudio de condiciones psicosomáticas que implican una

interacción entre el cuerpo y la mente, cuando se supone que esta última es un fenómeno metafísico. La idea de que el espacio somático está cerrado tiene un importante punto de partida en la conocida aseveración de T. H. Huxley, el gran biólogo inglés del siglo XIX, según la cual la conciencia es un efecto colateral del funcionamiento del cuerpo que carece de poder para influir en los procesos somáticos. Según Huxley, los humanos son autómatas conscientes. La conciencia sería un epifenómeno similar al silbato de una locomotora o a la sombra que sigue al paseante en un día soleado, fenómenos que no influyen en el movimiento del tren o en el andar del que camina. Esta manera de entender la conciencia se encuentra todavía muy extendida entre los neurocientíficos. A esta idea se agrega en muchas ocasiones una explicación simplista y reduccionista del problema de la conciencia: se trataría de un fenómeno o proceso que ocurre totalmente dentro del cerebro y que no tendría ninguna base extrabiológica.

Yo critiqué esta interpretación cuando desarrollé la idea de que la conciencia incluye una especie de prótesis simbólica que prolonga, en los espacios culturales, algunas funciones de las redes neuronales. Esta prótesis, que he definido como un exocerebro, está compuesta principalmente por el habla, el arte, la música, las memorias artificiales y diversas estructuras simbólicas. Algo muy importante en mi teoría radica en la afirmación de que los elementos exocerebrales de la conciencia tienen un poder causal y son capaces de modificar y modular la operación y las funciones de las redes neuronales. Estos circuitos exocerebrales no son instancias metafísicas y no se encuentran fuera de la clausura causal en la que los científicos circunscriben, con razón, sus explicaciones.

Para sustentar esta afirmación he emprendido la exploración del extraño y fascinante proceso que los médicos denominan efecto placebo. Este efecto es una realidad, y por lo tanto se trata de una prueba de que las estructuras simbólicas arraigadas en la cultura son capaces de influir en las funciones cerebrales por medio de la conciencia. Los estudios científicos muestran que el uso de sustancias farmacológicamente inocuas o la práctica de operaciones simuladas tienen repercusiones somáticas comprobables. La clave del efecto placebo se halla en el hecho de que el paciente cree firmemente que el remedio que aplica un médico (o un chamán) es eficiente. Se ha mostrado que el placebo produce efectos fisiológicos observables, como cambios en la presión, el ritmo cardíaco, la actividad

gástrica, y deja señales incluso en las redes neuronales. Los placebos han sido usados con éxito en el tratamiento del dolor, la ansiedad, la depresión, las úlceras, las enfermedades de la piel, la artritis reumatoide, el asma, la presión sanguínea alta, las enfermedades autoinmunes e incluso el párkinson. Menos estudiado es el efecto nocebo, el reverso del placebo.

Para iniciar mis reflexiones, quiero explorar un antecedente histórico medieval muy significativo del uso medicinal del efecto placebo. Se trata del texto de un médico sirio que vivió aproximadamente entre los años 830 y 910 d. C. Este médico medieval, por supuesto, no usó el término «placebo». Esta palabra tiene su origen en un error cometido por san Jerónimo al traducir del hebreo al latín el noveno versículo del salmo 114 (116). En lugar de traducir «Caminaré en presencia de Yahvé por la tierra de los vivos», escribió: «Placeré al Señor en la región de los vivos» (*Placebo Domino in regione vivorum*). A partir de esta equivocación sucedió que, durante la Edad Media, los profesionales encargados por las familias de llorar a sus muertos con frecuencia iniciaban el lamento artificial con el versículo 9 del salmo 114. Aquí el llanto «artificial» sustituía al «verdadero». Era frecuente que los sicofantes alquilados para abrir las Vísperas de Difuntos con su llanto artificial iniciasen los lamentos recitando ese mismo verso en latín. Por ello estos profesionales del lamento fueron llamados placebos, y así la palabra adquirió la connotación de adulador. Posteriormente el término se usó para referirse al medicamento falso que se recetaba para complacer al enfermo. El placebo acabó aludiendo a algo «artificial» que se cree «verdadero». Lo esencial es la creencia o la fe en un acto que se inscribe en el ritual oficiado por un brujo, un sacerdote o un médico, quienes con el poder de la palabra y de la simulación producen efectos curativos y placenteros.

Así pues, desde su mismo origen, la idea de placebo se liga a un ritual que tiene por objetivo disminuir el displacer ocasionado por las desgracias ocurridas, como la muerte de un ser querido, o las provocadas por la enfermedad. El ritual trata de dispensar placer y alivio a las personas dañadas o maltratadas. Así que el error de san Jerónimo nos conecta con la necesidad de proporcionar placer, de complacer, a quienes sufren. El efecto placebo está inscrito en los rituales del placer, apoyados en el inmenso poder de la palabra. Pero, a diferencia de lo que quería san Jerónimo, no buscan complacer a la divinidad sino aliviar a los humanos de los males que

los aquejan. A Freud le hubiera gustado pensar que los rituales del placebo se encuentran ligados al «principio del placer», a los procesos anímicos que empujan a las personas a buscar una satisfacción agradable, procesos solo reprimidos por el «principio de realidad», que para evitar peligros logra que se aplace la satisfacción o incluso que se renuncie a ella.²

2. LAS LIGADURAS DE QUSTA IBN LUQA

El médico sirio del siglo IX al que quiero referirme se llamó Qusta ibn Luqa al Ba'labakki, que traducido quiere decir Constantino hijo de Lucas nativo de Baalbek. La ciudad siria donde nació hacia 830 fue conocida como Heliópolis durante la época romana. Hoy se encuentra en el Líbano, a unos ochenta kilómetros al este de Beirut y unos cincuenta y cinco kilómetros al norte de Damasco. Qusta ibn Luqa, de origen griego, fue un cristiano melquita cuyos escritos en árabe fueron muy importantes en la transmisión de la sabiduría griega al mundo árabe.³ Vivió mucho tiempo en Bagdad y murió en Armenia en torno del 910. La influencia en Occidente de este médico, filósofo, astrónomo, músico y matemático árabe se puede comprobar por el hecho de que uno de sus libros, *Sobre la diferencia entre el espíritu y el alma* (o el pneuma y la psique), fue uno de los pocos textos no escritos por Aristóteles que fueron incluidos en la lista de lecturas que elaboraron los maestros de la facultad de artes en París en 1254 como parte del estudio de la filosofía natural. Tradujo muchas obras científicas griegas al árabe y fue autor de decenas de obras.⁴

Siendo cristiano, Luqa escribió una guía médica para los peregrinos a La Meca (*Risa-la Fi- Tadbir Safar Al-h. ag ģ ģ*), donde recomienda la mejor dieta para el viajero, las posibles enfermedades que lo pueden afectar por el camino y la manera de tratarlas. Se refiere a la fatiga, el dolor de oído, enfermedades de los bronquios, la drancunculiasis ocasionada por la lombriz de Guinea, las picaduras de insectos y problemas ocasionados por el polvo.

Qusta ibn Luqa escribió en árabe un breve tratado, que solo ha sobrevivido en su traducción latina, titulado *Ligaduras físicas, o sobre encantamientos, conjuros y colgantes en el cuello* [*De phisicis ligaturi (De incantatione adiuratione colli suspensione)*].⁵ No se conoce ningún manuscrito del original árabe. Hay quince manuscritos de la versión latina, los más antiguos del siglo XII. En forma impresa circuló como parte de las obras compiladas de otros autores, como Constantino el Africano, Galeno, Arnau de Vilanova y Cornelius Agrippa. Generalmente aparecía con el

título *De incantatione*. Es el primer tratado médico conocido que reconoce el efecto placebo, y que está abierto a nuevas perspectivas en el tratamiento de las enfermedades mentales. Este texto tiene la forma de una carta a su hijo, quien le ha preguntado si son efectivos los encantamientos, conjuros y colgijes, y si hay explicaciones para ello en los textos griegos, tal como sí las hay en los libros de los indios.

Qusta debió conocer bien la medicina hindú, que influyó mucho en las prácticas médicas del Bagdad de su época. Desde luego, en la India se practicaba una medicina de gran complejidad y sofisticación, y para nada se reducía al uso de conjuros y amuletos. Incluso hubo médicos hindús que ejercieron en Bagdad, y textos clásicos de la medicina hindú fueron traducidos al árabe. Manka, un célebre médico hindú que vivía en la corte de Harún al-Rashid, tradujo textos médicos del sánscrito.

Su texto *Ligaduras físicas* comienza por establecer que los antiguos «parecen estar de acuerdo en que la complexión del cuerpo corresponde al poder del alma» y que si esta se encuentra en equilibrio «la acción del cuerpo será igualmente perfecta». Pero advierte que si la complexión del cuerpo está en desequilibrio, el alma también será imperfecta. «Así», escribe Qusta, «se ve casi siempre que son imperfectos los movimientos del alma en los niños, los viejos y las mujeres», lo que confirma que «la complexión de su cuerpo es también imperfecta, y esta imperfección afecta también a los habitantes de regiones destempladas, como podemos esperarlo con el calor en Etiopía y el frío en Escocia».⁶ De aquí concluye que si alguien tiene confianza en un encantamiento, ello le ayudará, pues la complexión del cuerpo sigue la del alma: «y ello se comprueba», dice, «por el hecho de que el miedo, la tristeza, la alegría y el estupor provocan en el cuerpo no solo un cambio de color, sino también otras maneras, como la diarrea, el estreñimiento o la debilidad extrema». Y agrega: «Más aún, yo he visto que estas cosas son causa de una alteración prolongada de la salud, especialmente en las alteraciones que dañan la mente.»

En este contexto, la referencia que hace Qusta a Platón es muy significativa. Dice que, según Platón, cuando el entendimiento es firme, aunque de manera natural no sea saludable, un objeto podrá ser útil gracias a las intenciones de la mente, por lo que si alguien tiene confianza en un encantamiento, ello puede ayudar, y en esa medida cualquier cosa, sea lo que sea, lo puede auxiliar. Se trata de una referencia al diálogo *Cármides*.

Allí Sócrates le receta a Cármides, joven atractivo e inteligente, un remedio para sus dolores de cabeza. Le dice que debe tomar cierta hierba (*pharmako*), pero que es necesario añadir unas palabras mágicas (*epode*), un conjuro que Sócrates aprendió de un médico tracio discípulo de Zalmoxis, un dios que les prometió la inmortalidad a sus devotos.⁷ Sócrates advierte a Cármides: «Se trata el alma valiéndose de ciertas palabras mágicas. Estas palabras mágicas son bellos discursos. Gracias a estos bellos discursos la sabiduría toma raíz en las almas, y, una vez arraigada y viva, nada más fácil que procurar la salud a la cabeza y a todo el cuerpo.»

Es la curación por la palabra y Qusta la invoca siguiendo el ejemplo de los médicos indios que creen firmemente que los conjuros y los encantos son útiles. Añade una referencia a Hipócrates comentado por Galeno: «Si uno dice que cuando los humores corporales cambian, el movimiento del alma cambia de igual manera, entonces no es equivocado decir que la mutación del alma modifica a esos mismos humores.» Así, el humor colérico incrementa el conocimiento y el intelecto; la melancolía induce continencia e inclinación por el estudio; la complexión sanguínea impulsa el deseo de hablar y caminar; la flema, en cambio, no afecta al alma. En consecuencia, Qusta cree que de este principio se deriva que, a la inversa, el médico ayudará al alma por medio de un encantamiento, un conjuro o un colgante en el cuello; y, como consecuencia, se curará también el cuerpo.

A continuación Qusta procede a dar varios ejemplos. El primero es el caso de un gran noble que se queja ante el médico sirio de estar sujeto a una ligadura que le sujeta el pene y le bloquea el placer sexual. Se refería a un hilo invisible que mediante brujería o un conjuro diabólico volvía impotentes a los hombres. El reto consiste en cambiar el pensamiento del noble e inducirlo a creer en un remedio. El noble se resiste tercamente a cambiar de opinión hasta que Qusta le lleva el *Libro de Cleopatra* para leerle un párrafo. Se refiere a la *Cosmética*, un libro atribuido a Cleopatra que se conoce solamente por citas, pues no ha sobrevivido el texto. «Leí el pasaje», dice el médico sirio, «donde dice que alguien así ligado debe frotarse por todo el cuerpo bilis de cuervo mezclada con aceite de sésamo. Al oír esto confió en las palabras del libro y lo hizo; tan pronto fue liberado de la ligadura creció su deseo por tener relaciones.»

Después pasa a los ejemplos de objetos colgados al cuello con fines curativos. Cita los casos en que Aristóteles habla del poder curativo de las

piedras suspendidas en el cuello, como la esmeralda, que protege contra la epilepsia, el zafiro, que ahuyenta la peste, y la caledonia (especie de ágata), que evita las pesadillas. El ónix, en contraste, colgado al cuello induce al pleito y provoca malos sueños.

Escribe que Galeno recomienda para los dolores de estómago y vientre colgarse una bolsa con los excrementos de un lobo que ha comido huesos de oveja. Dioscórides tiene un remedio para la epilepsia: cuando la luna está creciendo, hay que cortar el vientre de una golondrina joven que está poniendo por primera vez para encontrar allí dos piedras, una de una sola tonalidad y la otra multicolor; hay que juntarlas en una bolsa de cuero de ternero para colgarla del cuello. También Dioscórides recomienda, para incrementar el deseo sexual, colgarse un ojo de cabra salvaje atado con raíces de mejorana dulce. Si el colguije es un dedo de niño abortado, la mujer no podrá concebir. Los médicos indios, por su parte, para impedir permanentemente el embarazo aconsejan verter excremento de elefante mezclado con aguamiel en la vulva de la mujer.

Qusta dice que no ha probado estos remedios aconsejados por los antiguos griegos ni tampoco los que recomienda la medicina védica de la India. Pero no niega su eficacia. Un médico griego del siglo II d. C., cuyas ideas seguramente conoció Qusta, Sorano de Éfeso, ya había abordado el tema del uso de amuletos mágicos y dijo que, aunque no tenían ningún efecto directo, no se oponía a su uso porque alegraban al enfermo. Sorano era conocido por su tratado de ginecología y aceptaba la magia y la superstición para que las pacientes se sintiesen en confianza con el médico, pero no creía que tuvieran ninguna influencia en el curso de la enfermedad. En cambio, la innovación de Qusta ibn Luqa consiste en afirmar que si el paciente tiene confianza en los objetos mágicos y los conjuros se produce un efecto saludable en su cuerpo. Es decir, reconoce el poder de la sugestión y la persuasión.

Qusta termina su texto inspirado en Platón, y afirma que un exceso de racionalidad de los médicos impide entender muchas enfermedades, pues se concentran solamente en el cuerpo y olvidan el conjunto, que incluye el psiquismo. «En algunos casos», dice, «ciertas sustancias tienen propiedades incomprensibles para la razón, porque su sutileza no es canalizada por los sentidos, debido a su gran profundidad.» Es decir que la sola razón no puede penetrar la esencia de ciertas cosas. Entre las cosas incomprensibles,

dice Qusta, tenemos la acción de un imán atrayendo hierro, el plomo que quiebra el diamante, cosa que no puede hacer el fierro, la soda (hidróxido de sodio) que no puede ser quemada por el fuego, y el hecho de que cierto pez al ser tocado ocasiona la pérdida del sentido (se refiere a la raya o la anguila eléctrica). Conocemos estos hechos por sus propiedades y no por las razones que permitirían entender su esencia. Así, dice Qusta, «las cosas colgadas del cuello ayudan por sus propiedades y no por su naturaleza interna», es decir, porque «fortalecen la mente». Es decir, se conocen sus propiedades ocultas solamente por sus efectos.

Estamos, creo yo, ante una explicación antigua de la manera en que ciertas prótesis simbólicas (los colgijes, los amuletos y los conjuros) se adaptan al funcionamiento de las redes neuronales, gracias a lo cual se producen efectos en el cuerpo. Es lo que caracteriza al efecto placebo. Es posible, como se ha dicho, que una buena parte de los tratamientos médicos usados antes de la era moderna fueran en realidad ejemplos del efecto placebo. Es por ello sintomático el hecho de que, aunque el texto de Qusta circuló ampliamente en Occidente, su explicación no se incluyó en los modelos de la medicina académica. No es sino hasta mediados del siglo xx cuando la medicina comienza a reconocer la importancia del efecto placebo.

Al leer las explicaciones de Qusta ibn Luqa puede asombrarnos su semejanza con los procesos descritos por los médicos actuales. Ciertamente, hoy los médicos no hablan de la influencia de los rituales mágicos y los conjuros en el equilibrio de los cuatro humores hipocráticos. Pero reconocen, en el efecto placebo, la influencia del entorno sociocultural y simbólico en los flujos de endorfinas y dopaminas en el cuerpo humano. Al mismo tiempo, hoy en las sociedades modernas todavía hay creencias populares en el poder saludable de toda clase de collares y pulseras magnéticas; es el caso de los famosos brazaletes Nikken, las pulseras Power Balance, los colgijes con metales o piedras que supuestamente contienen vibraciones de energía con propiedades curativas, como los collares de titanio o los que emiten iones negativos. Hay que agregar los escapularios devocionales con funciones de talismán o amuleto, tan comunes en la imaginería católica popular. Sean píldoras inocuas o collares energéticos, se trata de prótesis con poder sanador que son manipuladas por médicos o merolicos gracias al poder de la palabra.

El texto de Qusta enfatiza la importancia que tiene la manipulación del

alma para curar el cuerpo. Trata de dar una explicación hipocrática y platónica de las prácticas médicas que provienen, nos dice, de la India. En realidad, se refirió más a prácticas mágicas populares que a las complejas tradiciones médicas hindús, que usaban tanto la herbolaria como la cirugía. Acepta que no puede comprender cómo la *naturaleza* de los amuletos colgantes ejerce una influencia en el cuerpo, pero afirma que entiende cómo sus *propiedades* influyen –a través de la mente– en los procesos corporales. El uso médico del efecto placebo, tal como se practica hoy, parte de la premisa de la naturaleza inocua de sustancias o prácticas, pero acepta que tienen efectos en el cuerpo que es necesario comprender y que, aparentemente, pasan a través de la conciencia.

3. LOS PODERES MÁGICOS

Los antropólogos han observado en muy diversas culturas prácticas de curación, acompañadas de rituales y conjuros, que en cierta medida parecen producir efectos benéficos. Estas prácticas no siempre están enmarcadas en elaboradas explicaciones, como es el caso de la medicina antigua (hipocrática y védica), pero sí forman parte de un complejo tejido de mitos y ritos. Me refiero al chamanismo, que, como hiciera Qusta ibn Luqa, suele invocar una manipulación del alma, que se ha separado del cuerpo del enfermo; en consecuencia es necesario desencadenar un proceso (con frecuencia un viaje) para recuperarla y volverla a unir al cuerpo. En cierto modo, el chamán se propone manipular el alma (o las almas) para lograr su retorno. Es significativo el hecho de que el chamanismo, como un ritual practicado por un curandero, o, como se dice en inglés, un *medicine-man*, es un fenómeno cultural muy extendido entre los que se solía llamar pueblos primitivos de todo el mundo; es también una importante tradición en los sectores marginales de las sociedades modernas adonde no llega la medicina científica.⁸

Qusta ibn Luqa recorrió extensamente los territorios del antiguo imperio de Bizancio; invitado por el rey, se estableció entre los años 860 y 870 en Armenia, donde murió a comienzos del siglo x. Sin duda debió de observar en la región bizantina las prácticas chamánicas tan características de los pueblos altaicos y siberianos que se extendieron entre los turcos, los mongoles y los tunguses. Sus observaciones sobre los talismanes y los amuletos revelan su interés por las tradiciones mágicas populares, entre las cuales destacó el chamanismo. El concepto mismo de chamanismo surgió del estudio de los curanderos tunguses, y la palabra *chamán*, en la lengua de este pueblo, se refiere a los magos y curanderos que adivinaban el futuro y sanaban a los enfermos. La idea central del chamanismo siberiano y altaico es la representación del mundo dividido en dos niveles conectados por un eje central por el cual el chamán puede transitar. Los chamanes son capaces de entrar en un trance para realizar un viaje y con ello curar al enfermo o predecir el futuro.

El chamanismo es una práctica ritual antigua, como lo muestran los anales de los emperadores de la dinastía Tang, que usan la palabra turca para referirse al chamán (*qam*) cuando aluden a las prácticas mágicas de los kirguises. Estos anales comenzaron a compilarse en el año 941. El gran médico y filósofo Avicena, un siglo después que Qusta ibn Luqa, describió un ritual chamánico turco. Ya en el siglo XVII en Turquía los ritos chamánicos adoptaron la forma típica que han observado los etnólogos en el Altái y en Siberia. Las descripciones clásicas fueron consignadas extensamente por los antropólogos rusos V. M. Mikhailovskii y S. M. Shirokogoroff, que refieren con gran detalle las actividades de los chamanes, sus creencias, su vestimenta, su parafernalia y sus éxtasis. El chamán es fundamentalmente un manipulador de los espíritus que en ocasiones penetran en las personas y les producen daños. El mismo chamán es invadido por espíritus, pero no permite ser poseído por ellos.⁹ En ocasiones el chamán va contando paso a paso su viaje en busca del alma que ha escapado del cuerpo del enfermo y ha sido secuestrada por malos espíritus. Durante el camino se enfrenta a poderes malignos que bajo la forma de animales tratan de bloquearle el paso. Pero hay otros animales que lo apoyan. El chamán va vestido como animal (reno o pájaro), lleva un bastón con mango en forma de cabeza de caballo y toca un tambor con dibujos que representan las dos regiones del universo y el eje que las conecta. En la época otomana había «babas» que viajaban por toda Turquía y, aun cuando predicaban la doctrina islamista, eran chamanes disfrazados que practicaban los antiguos rituales mágicos.¹⁰ A partir de los estudios clásicos del chamanismo en Siberia y en el Altái, el concepto ha sido extensamente usado para describir fenómenos similares en muchas otras regiones del mundo.

Un aspecto esencial de estos rituales es el trance extático en el que entran los chamanes. Sin duda sería muy interesante explorar las repercusiones neuronales que genera el éxtasis chamánico, pero es un campo poco explorado. Me interesa aquí más explorar la relación entre el ritual chamánico y el cuerpo del enfermo. ¿Hay efectos somáticos de una actividad que es esencialmente simbólica? ¿Son eficaces los procedimientos chamánicos para lograr una sanación? El antropólogo Claude Lévi-Strauss planteó en 1948 preguntas similares cuando se ocupó de un canto sanador de los indios kuna de Panamá. Al comparar el ritual chamánico con el

psicoanálisis, Lévi-Strauss llegó a la conclusión de que en ambas técnicas hay una «eficacia simbólica»; es decir, hay una transformación orgánica lograda gracias a que el enfermo «vive intensamente un mito». La eficacia simbólica radica en la inducción de cambios orgánicos entre estructuras formalmente homólogas pero construidas con materiales diferentes (como el pensamiento reflexivo, el psiquismo y los procesos fisiológicos).¹¹ Ello significa que una actividad simbólica externa provoca en el cuerpo cambios orgánicos. En el chamanismo el enfermo vive un mito social tradicional; y en el psicoanálisis el mito procede de la historia individual. En el primer caso el chamán es el que habla; en el psicoanálisis el médico calla y el enfermo habla.

Un año antes, en 1948, el antropólogo italiano Ernesto de Martino se enfrentó al mismo problema.¹² ¿Son reales los poderes mágicos de los chamanes? Para este antropólogo, desde su perspectiva historicista, los rituales chamánicos se inscriben en un mundo mágico primitivo, una realidad muy distinta a la nuestra. Así, el chamanismo opera con poderes reales, que se pueden comprobar con lo que Ernesto de Martino llama el «escándalo de la investigación» sobre las capacidades de clarividencia, lectura del pensamiento, comunicación a distancia y métodos paranormales de conocimiento propios de los chamanes tunguses de Siberia. Ellos viven en una época mágica diferente a la civilización occidental. Para ilustrar su interpretación de la realidad de dos mundos, el mágico primitivo y el moderno civilizado, De Martino acudió además a la información de un misionero anglicano que había trabajado en el Chaco paraguayo, Wilfred Barbrooke Grubb. El ejemplo se refiere a un indígena de la etnia lengua (los enxet-enlhet) de Paraguay que acusó a Grubb de haber robado unas calabazas de su huerto, basado en el hecho de haberlo visto en sueños cuando las hurtó. El misionero no había cometido el robo, por supuesto, y desde el punto de vista occidental se trata de un hecho irreal, dice De Martino; pero concluye que no es así, por extraño que pueda parecer, pues el indígena se encuentra en la época histórica donde lo real culturalmente significativo incluye a la conciencia onírica como algo vivido.¹³ Grubb, en el contexto del mundo mágico, vivió realmente una existencia en el sueño de otra persona, aunque él no lo reconociese.

Ernesto de Martino era discípulo de Benedetto Croce. Pero al maestro no le gustó la idea de la existencia de dos mundos, nuestro mundo occidental y,

por otro lado, el mundo primitivo mágico donde los poderes paranormales chamánicos son reales. Croce critica la peculiar visión de «realidad» del antropólogo, quien sostiene que tanto el indígena soñante como el misionero afirman una realidad, cada uno la suya.¹⁴ Croce no acepta que las categorías propias de la interpretación histórica que usan los etnólogos no se puedan aplicar a las épocas primitivas. No se puede negar la perpetuidad de las categorías, que son como el «motor inmóvil» del que habló Aristóteles, dice Croce.

Mircea Eliade, que en esa época estaba preparando su famoso libro sobre el chamanismo, también publicó un comentario sobre *Il mondo magico*.¹⁵ Observa que De Martino, como idealista que sigue la corriente de Benedetto Croce, está convencido de que cada mundo es una creación espiritual de los hombres. Reseña además otra de las ideas centrales del libro del antropólogo italiano: en el mundo primitivo no se ha desarrollado la unidad de la persona, su individualidad, a diferencia de lo que ha ocurrido en Occidente. De allí que los primitivos vivan dramáticamente la amenaza de perder su ser y por ello encontramos esa idea recurrente del chamanismo: la búsqueda del alma perdida. Para los primitivos el mundo nunca está *dado* y el ser no está nunca *garantizado*, la naturaleza se encuentra culturalmente condicionada y por lo tanto los poderes paranormales son reales aunque en nuestra realidad occidental moderna no sean auténticos. De Martino, señala Eliade, se mantiene en la más pura perspectiva del idealismo historicista: el mundo no está *dado* sino que está *hecho* por los hombres. Eliade tiene una interpretación completamente diferente; para él los fenómenos paranormales no están únicamente encerrados históricamente en el mundo mágico primitivo, sino que también aparecen en la época moderna, sin que por ello sean una regresión histórica inauténtica. Eliade creía que lo sagrado es una vivencia trascendente que se puede experimentar realmente. Según Eliade es necesario considerar que hay «otro mundo accesible, en principio, a *todos* y en *cualquier momento histórico*». Esta afirmación nos conecta con la tesis junguiana de los arquetipos.

Quiero regresar a la historia del misionero y el indígena del Chaco que supuestamente viven en dos realidades diferentes. En realidad Ernesto de Martino cambió los hechos narrados por Grubb, para adaptarlos a su tesis, y no contó la historia completa. Sospecho que no tomó la historia

directamente del libro de Grubb. En su libro el misionero anglicano explica que el indígena soñador llegó desde un lugar muy lejano. Cuando Grubb le explicó que hacía mucho que no iba a su pueblo y que no podía haberle robado las calabazas, el indígena le dijo que lo había visto en sueños robando tres calabazas en su jardín. Al discutir con el misionero *admitió francamente que, en efecto, no le había robado las calabazas*. Sin embargo, el indígena exigió que se las pagase, pues –dijo– si hubiese estado allí se las habría robado, ya que lo que vio en su sueño fue el alma del misionero que estaba en su jardín, lo que demostraba que tenía la voluntad o la intención de robar y que el misionero lo hubiera hecho si su cuerpo hubiese estado allí. La creencia de los indígenas consistía, en realidad, en pensar que el alma, percibida gracias al sueño, revelaba la intención del cuerpo. Cuando tiempo después otro indígena, de nombre Puit, intentó asesinar a Grubb de un flechazo, fue ejecutado por los habitantes de su pueblo debido *a su intención*, aun cuando no lo hubiese logrado. Lo que subraya Grubb es la enorme importancia que los indígenas le dan a la intención, por encima de los mismos actos. Así pues, no hubo un misionero que en un mundo robó las calabazas y en otro mundo no lo hizo.¹⁶

Mucho antes que por Ernesto de Martino, el ejemplo del indígena del Chaco narrado por el misionero Grubb había sido utilizado por Lucien Lévy-Bruhl para describir a las que llamó «sociedades inferiores», caracterizadas por su aversión al razonamiento y sus inclinaciones prelógicas y místicas. Creo que esta es la fuente de la explicación de Ernesto de Martino, quien probablemente no consultó el texto de Grubb. En su famoso libro de 1922 *La mentalidad primitiva* Lévy-Bruhl expuso la idea según la cual para estos pueblos inferiores el mundo visible y el mundo invisible forman un todo.¹⁷ Para probar su tesis usa el caso del indígena que soñó que el misionero Grubb le había robado sus calabazas. Cuenta la historia completa y comprende que el problema radica en que para el indio los sueños revelan las intenciones y no forzosamente los actos. Pero Lévy-Bruhl hace a un lado el tema de las intenciones, para destacar en cambio su definición de la mentalidad primitiva como prelógica. Lévy-Bruhl supone que el indio acepta lo que los escolásticos llamaron la «multipresencia», es decir, la ubicación de un mismo ser en lugares diferentes. Esta afirmación no se sustenta en los hechos que narra Grubb: el indio lengua no cree que Grubb estuvo en dos lugares al mismo tiempo.

Cree que lo que ha visto en su sueño revela las intenciones de Grubb. Sin embargo, con ello Lévy-Bruhl quiere ilustrar su idea de que los primitivos carecen de lógica. Supone, erróneamente, que el indígena que sueña está aceptando implícitamente algo verdaderamente contradictorio. Lo equipara a la creencia en muchas sociedades inferiores en que la persona recién muerta que se aloja en su tumba al mismo tiempo deambula por los lugares donde vivió. Lévy-Bruhl compara este pensamiento prelógico con la interpretación religiosa de la eucaristía, donde el cuerpo de Cristo se halla presente en diferentes lugares al mismo tiempo. En esta multilocación el cuerpo sigue siendo uno solo pero su presencia se multiplica. ¿Se habrá preguntado Lévy-Bruhl si la creencia en la multilocación del cuerpo de Cristo es una forma prelógica de pensamiento? Sí le inquietó el problema, pues en sus *Carnets*, muchos años después, vuelve sobre la historia de Grubb. Allí reconoce que haber definido al indio lengua como dotado de una mentalidad prelógica que caía en contradicciones no era correcto. Comprende ahora que era una imposibilidad *física* que Grubb estuviera en dos lugares diferentes al mismo tiempo, pero que no había una contradicción lógica. Y señala que los cristianos creen en la multipresencia de Dios (de hecho, la omnipresencia). Debo decir, para ser justo, que en 1938, hacia el final de su vida, Lévy-Bruhl hizo en sus *Carnets* una autocrítica de su idea de la mentalidad primitiva: «corrijamos expresamente lo que yo creía exacto en 1910: no hay una mentalidad primitiva que se distinga de la otra por *dos* caracteres que les son propios (mística y prelógica). Hay una mentalidad mística más marcada y más fácilmente observable entre los “primitivos” que en nuestras sociedades, pero que está presente en todo espíritu humano». ¹⁸

El último capítulo de *La mentalidad primitiva* está dedicado a la relación de los primitivos con los médicos europeos. Lo que intriga a Lévy-Bruhl es el hecho, relatado en muchas ocasiones por misioneros y médicos, de que los indígenas al ser curados, en lugar de estar agradecidos, le exigen un pago al europeo por haberse dejado tratar. Se da cuenta de que los primitivos aceptan más fácilmente ser medicados si tienen confianza en el médico, pues suponen que las medicinas están dotadas de propiedades místicas. Cuenta que a los fang de Gabón, en África, les asombra que los europeos no usen cantos, exorcismos ni conjuros de ninguna clase. En un caso el enfermo sabía un poco de francés y no le parecía asombroso que el

remedio no sirviese para nada sin ir acompañado de algún ritual. El médico que lo atendía casi no había abierto la boca, y lo poco que había dicho era «¡traga, negro cochino!». En consecuencia, el fármaco no tuvo efecto alguno. En cambio, un médico que mientras atendía o curaba siempre cantaba aires alegres, gozaba de la confianza de los enfermos. Lévy-Bruhl explica la idea de que el médico debe pagar al enfermo que se deja curar por la creencia de que, al ser tratados por un europeo, se ha provocado que el indígena pierda el apoyo de las fuerzas invisibles que normalmente lo ayudan, de manera que el extranjero tiene ahora una responsabilidad que lo obliga a ayudarlo. No me parece una explicación convincente. Pero lo interesante aquí es que Lévy-Bruhl no se preocupa por investigar si la medicina practicada con rituales tiene un efecto real y positivo que se suma a los efectos de los fármacos. Este problema es el que inquietó a Claude Lévi-Strauss y, de alguna manera, a Ernesto de Martino.

No quiero dejar la impresión de que el chamanismo es una práctica localizada solamente en las sociedades llamadas primitivas. Ya he mencionado el uso con fines sanadores en sociedades urbanas modernas de brazaletes o collares magnéticos y de colgijes con metales que contienen vibraciones curativas. Pero una de las prácticas chamánicas más extendidas en sociedades industriales y urbanas ocurre en las numerosas iglesias espiritualistas, muy extendidas en Estados Unidos y en el Reino Unido. El ritual en estas iglesias tiene como eje central la actuación de un médium capaz de comunicarse con el mundo invisible de los espíritus. Con frecuencia el médium es un sanador que contribuye a aliviar males con ayuda de los espíritus. Los espiritualistas no suelen usar el término «chamán» para referirse a sus médiums, pero un antropólogo podría reconocer fácilmente las prácticas típicas del chamanismo en estas personas. Hay una investigación muy reveladora hecha por un profesor de la Universidad de Edimburgo, que es también un médium en la iglesia espiritualista de su ciudad, en Escocia. El libro que escribió es fascinante, pues combina la objetividad de un etnólogo con la fe en sus poderes como médium. Durante su trabajo de investigación fue entrenado como médium, paso a paso, hasta que alcanzó la capacidad, en las sesiones de los círculos de meditación de la iglesia, de entrar en comunicación con espíritus-guías de personas fallecidas que lo acompañaron en sus viajes por el mundo invisible. Lo interesante es que este profesor, David Gordon Wilson, acepta

plenamente que se trata de una práctica chamánica similar a las que han descrito los antropólogos y que el médium es un chamán que tiene poderes sanadores.¹⁹ Así que Lévi-Strauss hubiese perfectamente podido agregar el ejemplo de estas congregaciones espiritualistas en sus estudios sobre la eficacia simbólica del psicoanálisis, similar a la que ocurre en las prácticas chamánicas.

4. UN VIAJE CHAMÁNICO EN BUSCA DEL ALMA PERDIDA

Cuando Lévi-Strauss analizó el ejemplo del canto que narra el ritual de un chamán kuna de Panamá, creyó encontrar allí los mismos procesos sanadores que ocurren en el psicoanálisis y que son efectivos. LéviStrauss asume que el canto era recitado por el chamán durante el ritual curativo. No es seguro que haya sido así, pero ciertamente es un canto que transmite con detalle una enseñanza antigua, un método de sanación. Examinar este ejemplo permite ver de cerca el curso de una sanación tal como es invocada por el chamán y que conocemos tanto en su versión en lengua kuna como en su versión pictográfica. La escritura pictórica consiste en una serie de dibujos que representan en forma simplificada los temas del canto sanador, titulado *Mu-Igala* (el camino de Muu), y que sirven de recurso mnemotécnico para apoyar la recitación.²⁰ A continuación haré un resumen muy apretado del viaje chamánico por el interior del cuerpo de la enferma.

El canto describe el malestar de una mujer preñada que tiene dificultades para parir y todo lo que hace el chamán (*nele*) para aliviarla. La mujer está sentada en una hamaca con las piernas abiertas y el chamán se sienta debajo de ella y enciende un brasero para quemar granos de cacao.²¹ El humo llena la cabaña en la que se encuentran y da fuerzas al curandero y a sus ayudantes, que son unos espíritus que encarnan en imágenes sagradas (*nuchu*). Estos asistentes son llamados «jefes de plantas» (*sailagan*) por su relación con hierbas medicinales y se preparan para enfrentar a Muu, definida como la «abuela principal», una fuerza que en la cosmología kuna es la causante de la formación del feto. Pero Muu es también la causa de la enfermedad que aqueja a la mujer, que con las piernas abiertas sangra y suelta flujos. La enferma dice que Muu quiere quedarse para siempre con su *nigapurbalele*, que es su espíritu vital, su alma. El chamán ha esculpido unas figurillas de madera (*nuchu*) en las que encarnan sus ayudantes. Ellos ahora son también chamanes (*nele*, plural *nelegan*).

A continuación los *nelegan* toman el camino de Muu, en la misma dirección que usa el miembro viril para entrar en la mujer. Caminan en fila, iluminan el sendero oscuro, en un viaje a través de la vagina hacia la matriz. Atraviesan la montaña baja, la montaña alta, la montaña puntiaguda y muchas más hasta llegar al remolino (*pirya*) de Muu. Allí hay caimanes y otras bestias horribles. En el silencio de la medianoche el viento se lleva unos troncos que obstruían el paso. El camino de Muu ahora va cuesta arriba, brilla en tonalidades doradas del color de las llamas. Muu y sus hijas los están esperando en su casa. Entonces Muu se pregunta: ¿qué quieren estos chamanes que suben por mi río? En la puerta del lugar de Muu vive una araña que no los deja pasar. En otra puerta hay bestias amenazadoras: un animal negro que sacude una cadena de hierro, un jaguar, un animal rojo, un gato. Muu advierte que habrá una lucha con el chamán. La casa de Muu está protegida por tejidos y cortinas que la rodean y que se convierten en oro. El chamán ordena cortar estos tejidos interiores y hace que el viento los tire y se los lleve. Ahora el camino está abierto. El alma (*nigapurbalele*) de la mujer encinta está allí quejándose de que Muu la tiene cautiva y que ha esperado mucho tiempo. Los chamanes dentro de la casa, que está manchada de sangre, echan humo con sus sombreros y obligan a Muu a rendirse. Toman el alma y los espíritus de la mujer y regresan por el mismo camino, marchando en fila, hasta llegar a la cabaña donde comenzó el viaje. Le dicen al alma: aquí frente a ti está en la hamaca tu cuerpo, y la colocan de nuevo en su lugar. La mujer se reanima y se buscan plantas medicinales para terminar de curarla. El chamán con las plantas se acerca a la mujer como un falo y se mueve como el órgano masculino posiblemente para desprender la placenta que no sale y limpiar los lugares interiores y secarlos. El canto describe más rituales para impedir que la mujer vuelva a perder el alma.²² Pero el flujo de sangre no ha sido detenido y se aplican más hierbas medicinales para interrumpirlo. Llega la partera, que examina los genitales de la enferma y los limpia. Por fin, con la ayuda del chamán la criatura es depositada sobre una hoja de *urwa*, que se pone roja por la sangre. El canto describe cómo un río dorado y otro plateado irrumpen y se llevan las enfermedades, los caimanes, los troncos de árbol y, finalmente, la «calabaza dorada» (posiblemente la placenta). Ahora soplan vientos del color del oro y de la plata.

No es muy probable que una mujer kuna en el trance de un parto difícil,

presa de fuertes dolores, pueda seguir con atención el complicado relato del canto chamánico (que he simplificado mucho en mi descripción). Pero seguramente, empapada en la cultura de su etnia, tiene nociones del papel del chamán y de las figuras talladas en madera balsa (*nuchu*), que son como un ejército de curanderos que debe penetrar en procesión por la vagina para buscar y capturar su alma perdida en las profundas entrañas de su cuerpo. Los etnólogos saben que en cada vivienda kuna solía haber una caja llena de *nuchus*, cada uno con su personalidad y funciones específicas. Las familias, si lo deciden, llevan sus figurillas a otras casas para ayudar a curar a un enfermo. Para que adquieran poder hay que animarlas con humo de cacao o de pimienta.²³ El antropólogo Michael Taussig piensa que Lévi-Strauss impone un orden en un texto originalmente desordenado y caótico. No cree que el canto invoque el control del caos para dar nacimiento a una nueva vida. Afirma que Lévi-Strauss en realidad realiza a su vez un ritual mágico con un halo científico. Desde luego, el texto del canto tiene algunas contradicciones, pero no podría decirse que es caótico.²⁴ A mí me interesa el hecho de que, como ya lo señalé, Lévi-Strauss enfatiza las semejanzas del ritual kuna con el procedimiento psicoanalítico. Para ello compara la «eficacia simbólica» del rito kuna con la «realización simbólica» del método usado por la psicoterapeuta suiza Marguerite Sechehaye para lograr la curación total y definitiva de una esquizofrénica. El estudio de Sechehaye se publicó en 1947, cuando Lévi-Strauss estaba escribiendo sobre el chamanismo de los kuna. El punto de partida de Sechehaye es su creencia de que el Yo de la enferma se ha desintegrado y es necesario un viaje para lograr su reconstrucción. El Yo se supone que ha sufrido una regresión a una fase mágica, a una etapa primitiva y animista, debido a los malos tratos y frustraciones que la esquizofrénica sufrió desde muy pequeña. La doctora Sechehaye se da cuenta de que debe viajar a esa época primitiva y fetal donde las palabras carecen de sentido; recurre entonces a los símbolos para ir guiando al Yo en su camino hacia su reconstrucción, como el chamán kuna que penetra por la vagina de la parturienta en busca de su alma perdida. Para ello la psicoanalista solo podía usar un lenguaje de signos simbólicos, gestos y movimientos para comunicarse con el Yo sumergido en la magia primigenia. Utiliza objetos y muñecos simbólicos, junto con un ritual cariñoso. Poco a poco la enferma va abandonando su conducta mágico-animista y artificialista, asegura Sechehaye, para adaptarse

socialmente: se restablecen los límites entre el Yo y la realidad, que se habían disuelto. Hay una «realización simbólica» que reintegra al Yo perdido, una verdadera marcha ascendente de retorno a la realidad. El relato de esta curación es fascinante y recuerda sin duda los procesos rituales chamánicos; al interés que despierta la descripción del método usado se agrega el valor testimonial del diario escrito por la misma enferma, que describe con detalle las etapas de su recuperación.²⁵

Yo creo reconocer, tanto en las manipulaciones rituales chamánicas como en los rituales psicoanalíticos que he descrito, un denominador común: el uso de símbolos, que encarnan en palabras, objetos y actuaciones, como una expresión de lo que hoy los médicos denominan efecto placebo; los conjuros, collares y encantamientos, sobre cuyas propiedades terapéuticas reflexionó Qusta ibn Luqa, cumplían una función similar a la puesta en escena del proceso chamánico, lo mismo que en la sanación por medio de la palabra de los psicoanalistas. Lévi-Strauss con razón mostró las similitudes entre estos rituales sanadores, pero no pensó que podían tener algo más en común: el efecto placebo. Ahora quiero abordar directamente el estudio de este efecto para explorar una dimensión reveladora de los vínculos entre los circuitos neuronales y el fenómeno (no el epifenómeno) de la conciencia humana.

5. NEUROLOGÍA DEL EFECTO PLACEBO

Lo primero que es necesario establecer es que los placebos no son sustancias inertes; son más bien un conjunto de palabras, rituales, símbolos y significados que producen efectos somáticos comprobables.²⁶ Los placebos son una especie de simulacro durante el cual se administran píldoras o inyecciones con sustancias inocuas, o se practican operaciones quirúrgicas fingidas, con la intención de aliviar al enfermo. El interés por el efecto placebo comenzó a extenderse a mediados del siglo xx. Uno de los pioneros fue el doctor Henry K. Beecher, que se inspiró en sus experiencias como médico durante la Segunda Guerra Mundial, en África e Italia, y en los experimentos que hizo con prisioneros en Alemania en los años cincuenta (que, por cierto, fueron muy criticados por razones éticas).²⁷

Es necesario distinguir el efecto placebo de otros efectos como el condicionamiento pavloviano. Tanto humanos como animales pueden ser condicionados, como el famoso perro de Pávlov, que salivaba con solo escuchar una campana que había oído cuando le daban un pedazo de carne jugosa. Por ejemplo, se ha experimentado con el reflejo pavloviano cuando se administró durante varios días un compuesto de píldoras azucaradas que además contenían un fármaco inmunodepresivo. Cuando después se administró la píldora sin el fármaco, los sujetos respondieron como si lo ingiriesen. Algo similar ocurrió durante un experimento con ratas. Un psicólogo les inyectó una mezcla de agua dulce con otra sustancia que resultó ser letal (sin el psicólogo haberlo previsto). Cuando se percató de su error y vio que las ratas comenzaban a morir, les inyectó solamente agua con sacarina, pero los animales siguieron muriendo.²⁸ En otro ejemplo, varias personas fueron inyectadas durante dos días seguidos con una sustancia que provoca un aumento de secreción hormonal y una disminución de cortisol. Al tercer día fueron inyectadas con una sustancia inocua, sin que los pacientes se enteraran del cambio: tuvieron los mismos efectos que experimentaron al comienzo.²⁹ Desde luego, también hay que distinguir –lo que no siempre es fácil– entre el efecto placebo y la curación natural o la remisión espontánea.

El efecto placebo requiere de una participación consciente del enfermo. Por ello, el mal funcionamiento de los centros prefrontales de control (como en el alzhéimer) frena la eficacia del placebo. La desactivación de la corteza frontal del cerebro reduce o anula el efecto placebo.³⁰ Así, el efecto placebo no obedece al mismo proceso pavloviano de sustitución de estímulos. El placebo no es el sonido de una campana que sustituye al estímulo de la jugosa carne. Las píldoras inocuas no son como la campanada del perro de Pávlov, un mero estímulo que sustituye a un fármaco. Se ha señalado que los reflejos condicionados son una reacción posterior, mientras que el efecto placebo es un proceso cognitivo consciente, el resultado de una expectación previa. Se ha dicho, criticando este tipo de explicaciones, que una expectativa subjetiva privada al tomar una pastilla inocua no puede producir un cambio público observable en la fisiología del cuerpo.³¹ Debido a esta objeción, se ha buscado una explicación alternativa basada en la «intencionalidad motora» propuesta por Maurice Merleau-Ponty y que describe la generación de efectos fisiológicos que no requieren de conceptos previos: el cuerpo tiene su entorno y lo entiende sin necesidad de usar las funciones simbólicas.

Otros, en la misma línea inspirada en MerleauPonty, destacan la importancia de lo que llaman «experiencia encarnada» (*embodied experience*), que es definida como la incorporación biológica de materiales procedentes del mundo social en que vivimos.³² Esta es una manera de eludir el problema de explicar cómo los conceptos y los símbolos efectivamente producen efectos fisiológicos. Esta evasión es necesaria solamente si se considera que es imposible que haya una «traducción» de símbolos culturales a señales biológicas. Hay que subrayar que la realidad comprobada del efecto placebo indica que, de alguna manera, esta traducción ocurre en los humanos. Más adelante abordaré este problema.

El antropólogo Daniel Moerman sugiere un simpático experimento mental para entender la sorpresa que provoca entre los médicos el efecto placebo. Imaginemos que fabricamos unas llaves de tuercas que son idénticas en todo a las verdaderas salvo por el hecho de que están diseñadas para que, al intentar usarlas, la tuerca queda suelta y la herramienta no puede darle vuelta: es inocua. Si colocamos al azar estas llaves inútiles en las cajas de herramientas de los mecánicos de un taller, sería muy desconcertante que estos instrumentos realmente desenroscasen las tuercas.

Es lo que sucede cuando se administra al azar una pastilla inocua y se comprueba que sí alivia a los enfermos. El experimento mental sirve para reflexionar sobre la influencia del modelo mecanicista en nuestro pensamiento.³³

El efecto placebo ha sido comprobado en las frecuentes pruebas clínicas de los laboratorios farmacéuticos que contrastan la eficacia de un fármaco con placebos. También se confirma en experimentos de laboratorio encaminados directamente a medir la intensidad del efecto placebo. En las pruebas realizadas por empresas farmacéuticas donde se trata de probar una nueva medicina, se suele avisar a los pacientes que, al azar, pueden tomar un placebo o el medicamento verdadero. En estos casos, como bajan las expectativas de los enfermos, los efectos suelen ser minimizados. Los investigadores que hacen estas consideraciones se preguntan: ¿por qué existe el efecto placebo? ¿Por qué ocurre una respuesta efectiva cuando el placebo es administrado durante una relación interpersonal que requiere de pacientes atentos y conscientes? Su respuesta, como hipótesis, es que el efecto placebo activa las funciones autocurativas propias del organismo. Creen que el efecto placebo es eficaz para *aliviar el malestar* más que para *curar o controlar la enfermedad*. En consecuencia, concluyen, el efecto placebo, ya que opera principalmente sobre el malestar y no sobre la enfermedad, no puede haber surgido gracias a la selección natural. Fue más bien un subproducto de la prolongada crianza de los niños y de la solidaridad social en las primeras comunidades humanas.³⁴

Fabrizio Benedetti, que se ha dedicado a investigar el efecto placebo, señala que ha sido claramente probada su eficacia en varias dolencias. Es especialmente notable el efecto analgésico de los placebos; es el ejemplo más estudiado y que se entiende mejor. El mal de párkinson también ha sido un excelente modelo para comprender las respuestas orgánicas a la administración de placebos. También se ha probado que hay respuestas inmunológicas y hormonales inducidas por placebos. Se cree que el 75 % del efecto de los fármacos usados para tratar desórdenes psiquiátricos son atribuibles al efecto placebo. Asimismo, el uso de placebos en los deportistas aumenta su resistencia y su fuerza.³⁵ Es particularmente reveladora la relación entre el efecto placebo y el uso de antidepresivos. Un estudio sobre las pruebas aplicadas en laboratorios farmacéuticos muestra que el efecto placebo es fundamental: resulta que no hay diferencias

significativas entre el placebo y el fármaco salvo en los casos de depresión muy seria.

Además, las pruebas de laboratorio mostraron que el nivel de serotonina no es determinante en la depresión, ya que algunos fármacos la estimulan y otros la inhiben y sin embargo producen efectos similares. Si los fármacos activos producen resultados en gran parte generados por el efecto placebo, ¿qué se puede hacer? Como los antidepresivos producen fuertes efectos colaterales (como disfunción sexual, insomnio, náuseas, diarrea e incremento de peso), ¿es una opción prescribir placebos? Ello implica basar el tratamiento en una mentira y por ello se dice que es éticamente reprobable.³⁶ Otro estudio señala la misma situación: los pacientes mejoran igualmente tanto si toman placebos como si toman antidepresivos. Pero se agrega que también mejoran con las denominadas «terapias exóticas» no científicas. La dificultad radica en que, subrayan, la mayor parte de los depresivos terminan en una remisión espontánea. Pero al usar placebos hay un problema que se agrega al conflicto ético: los pacientes acaban percatándose de que no están tomando el fármaco efectivo ya que no sienten los efectos colaterales que les provocaban los antidepresivos.³⁷ Acaso haya un falso problema ético, pues el uso de placebos es una mentira y una simulación que se vuelve verdad. Pero la generalización masiva de placebos se toparía ciertamente con problemas prácticos muy difíciles de resolver. Además, el efecto placebo es un escudo que ayuda a que prosperen los tratamientos pseudocientíficos como la homeopatía, que es un gran negocio. Miles de personas que escogen alternativas inservibles para curarse, por ejemplo, de cáncer mueren a causa del error que cometen.

Menos investigado ha sido el efecto contrario al placebo, el nocebo, una consecuencia negativa producida por sustancias inocuas. Aquí también la creencia de que una sustancia genera malestar es lo que dispara los efectos nocivos. Se ha comprobado el efecto nocebo como resultado colateral aumentado de algún fármaco, como por ejemplo en el tratamiento con finasterida a pacientes con hiperplasia prostática benigna. A un grupo de enfermos en tratamiento se le explicaron con mucho detalle los posibles efectos colaterales de este fármaco, como la disfunción eréctil, pérdida de libido y problemas de eyaculación. A un segundo grupo no se le dio

ninguna explicación. El resultado de esta prueba fue que el 30,9 % del primer grupo reportó sufrir de disfunción eréctil, mientras que en el segundo grupo (no informado) solo lo reportó el 9,6 %. El 23,6 % del primer grupo sufrió disminución de la libido, mientras que en el segundo grupo solamente fue el 7,7 %. Algo similar ocurrió con los problemas de eyaculación, aunque la diferencia fue menos marcada (16,3 % y 5,7 %).³⁸ Podemos suponer que en el caso del efecto nocebo operan procesos muy similares a los que se observan en el caso del placebo.

El efecto placebo permite explorar un tema fundamental: los mecanismos que permiten traducir los *símbolos* culturales a *señales* neuronales. Se ha descubierto que los placebos que eliminan o alivian el dolor conducen sus efectos de manera similar a los fármacos opioides como la morfina. En 1978 se descubrió que los efectos analgésicos de una sustancia inerte eran bloqueados por el naloxone, un antagonista de los opioides usado para bloquear los efectos de sobredosis de morfina. Ello permitió entender que estos efectos analgésicos ocurren por mediación del sistema opioide endógeno gracias a la activación de sus receptores μ (mu).³⁹ Después se comprobó que la analgesia no solamente es mediada por el sistema opioide, sino también a través de otras vías que en el cerebro modulan la percepción del dolor. Por ejemplo, se ha estudiado la activación de un sistema inhibitorio endorfinérgico; cuando el placebo produce efectos en casos de lumbalgia se comprueba que el nivel de endorfina se ha elevado considerablemente.⁴⁰ Los sistemas moduladores internos del dolor son influidos por los estados emocionales y por las contingencias ambientales. Un ejemplo revelador es la analgesia provocada por la tensión extrema o estrés, como en el caso de los deportistas o los soldados heridos durante la competencia o la batalla, que han revelado no sentir el dolor. Entre paréntesis, quiero decir que el tratamiento del dolor no es un asunto marginal. En Estados Unidos el 10 % de la población sufre dolor crónico y muchos más lo sufren esporádicamente. Por ello hay un elevadísimo consumo de opioides, aunque hay un efecto colateral maligno: las sobredosis matan a miles de personas al año. No se ha estudiado la ayuda que podrían prestar los placebos ante la crisis debida a sobredosis de opioides recetados por los médicos (muchos decesos son provocados por drogas ilícitas).

El hecho es que la administración de placebos para el dolor activa los

receptores opioides μ del sistema endógeno. Pero, como dije, la analgesia también es inducida por otros caminos, y es modulada por las áreas cognitivas del cerebro, como los centros del lenguaje en el hemisferio dominante y otros centros de ejecución funcional en el hemisferio no dominante. Cuando la analgesia es más duradera, suelen estar implicadas las áreas emocionales ubicadas en la corteza temporal y parahipocámpica.⁴¹ Esta observación es muy importante, pues permite ver que los efectos del placebo están muy ligados a la conciencia y a las emociones. Sin una participación consciente y emocionalmente comprometida del enfermo, el efecto placebo se anula. Las sustancias inertes administradas cuando el paciente está anestesiado o inconsciente no producen ningún efecto.

Fabrizio Benedetti ha señalado claramente que los placebos activan las mismas vías bioquímicas que son activadas por los fármacos. Los humanos están dotados de sistemas endógenos que pueden ser activados por expectativas positivas inducidas verbalmente, por rituales terapéuticos y por diversos símbolos sanadores. Observa que el placebo usado como analgésico activa por lo menos dos sistemas endógenos: el opioide y el cannabinoide. Además, el sistema colecistoquinérgico influye en el circuito opioide y con ello estimula efectos placebo de diferentes magnitudes. Por su parte, el placebo administrado a enfermos de párkinson estimula la liberación de dopamina en el núcleo estriado del cerebro. Cuando es inyectada, la morfina se liga a los receptores opioides y por ello inhibe la transmisión del dolor; pero al mismo tiempo el ritual de su administración activa esos mismos receptores. Lo mismo sucede cuando se administra un fármaco dopaminérgico contra el mal de párkinson: se estimulan los receptores de dopamina; pero el mismo ritual es capaz de estimularlos también.⁴²

Se sabe que cuanto más espectacular es el simulacro más poderoso es el efecto. Los simulacros de cirugía son mucho más efectivos que la administración de píldoras de azúcar. Una investigación ha seguido el caso de 200 pacientes con una arteria severamente bloqueada que fueron operados para insertarles un stent. Al azar, a unos se les implantó el dispositivo y a otros no. La operación consiste en insertar un catéter por la arteria desde la ingle o el brazo para llegar al punto obstruido, inflar allí un globo para ensanchar el conducto e introducir un tubo metálico para mantener el flujo de sangre. En los casos en que se trató de un simulacro, el

catéter fue realmente insertado hasta el bloqueo pero solo se fingió hacer el resto de la operación. Lo notable es que seis semanas después no se detectó ninguna diferencia entre los dos grupos de operados: todos los pacientes sufrieron menos dolor y pasaron bien las pruebas de ejercicio en las cintas rodantes para correr.⁴³ Se vio que la aterosclerosis es parte de una enfermedad sistémica más amplia y que el remedio no consiste solamente en destapar arterias como si fueran las cañerías obstruidas del baño. Pero lo que a mí me interesa destacar es el hecho de que cuanto más espectacular e invasivo es el ritual, mayor es el efecto placebo.

Los ejemplos sobre la manera en que el placebo influye en el curso de las dolencias son múltiples y variados.⁴⁴ Los casos que he resumido son suficientes para comprender que estamos ante un proceso de traducción de los significados y símbolos implicados en el ritual sanador a señales neuroquímicas en los sistemas endógenos del sistema nervioso central. Las palabras y la parafernalia que usan los médicos son traducidas a flujos de las endorfinas, opioides, dopaminas y serotoninas que influyen en el funcionamiento del cerebro y en los procesos biológicos del organismo. Creo que el efecto placebo revela la presencia de un sistema de sustitución simbólica de elementos que faltan en el proceso biológico. Este proceso biológico de curación o de alivio se completa bien (o se completa mejor) con el apoyo de prótesis simbólicas. Las prótesis están formadas por los rituales sanadores que, como lo prueba el efecto placebo, tienen una importante influencia en el curso de un padecimiento, aun en ausencia del apoyo de fármacos activos. Ello solo puede ocurrir mediante la traducción de símbolos culturales a señales neuroquímicas. Falta investigar con precisión la manera en que ocurre esta traducción de símbolos a códigos químicos.

Se trata de un mecanismo de retroalimentación, pues los resultados positivos del efecto placebo estimulan la ampliación y consolidación de los rituales sanadores. Pero hay que agregar otra dimensión neuronal a los procesos endógenos estimulados por los placebos. La retroalimentación es un circuito que atraviesa el sistema emocional de las personas. Como se ha visto, el efecto sanador del placebo solamente opera por medio del sistema cognitivo: si no funcionan los mecanismos conscientes no hay alivio. El proceso es mediado por la conciencia. Pero tampoco hay efecto placebo sin una respuesta emocional de bienestar o calma. Y las emociones son, como

ha explicado bien Antonio Damasio, series complicadas de respuestas químicas y neuronales, que implican la secreción de hormonas (como el cortisol), péptidos (como la β endorfina), monoaminas, norepinefrina, serotonina y dopamina.⁴⁵

Estamos ante un circuito que comienza con las palabras del médico, que es una especie de chamán de bata blanca, las cuales estimulan mecanismos neuroquímicos de alivio. A su vez, estos mecanismos de alivio generan estados emocionales de bienestar, placer y calma que, además, estimulan expresiones y sentimientos de confianza hacia los médicos. Esta confianza apoya a los médicos cuando ejecutan sus rituales de sanación. Como puede comprenderse, el circuito atraviesa por las instancias simbólicas conscientes y por los sistemas neuroquímicos del cerebro. Hay una estrecha red de interconexiones que une los espacios biológicos con los culturales.

6. SOBRE AMULETOS ELECTRÓNICOS Y CATARSIS

Los rituales médicos forman parte de un conjunto que incluye instrumentos, máquinas, objetos y símbolos que confluyen en torno nuestro y que nos sirven de apoyo, nos alivian, nos informan, nos comunican, nos divierten y nos ayudan en nuestro trabajo. Una buena parte de estas prótesis, ceremonias y símbolos que se arremolinan cotidianamente a nuestro alrededor contribuyen a generar calma, bienestar y placer. En gran medida son una extensión de la conciencia. Son la parte artificial de nuestra conciencia. También nos divierten y nos hacen pensar. Generan un entorno agradable, tranquilo e incluso entretenido, pero con algunas fisuras que provocan malestar, tensión, disgusto, enojo, miedo o tristeza. Hay una extraordinaria prótesis que se ha extendido enormemente y que nos conecta con el entorno: el teléfono celular inteligente, que se ha convertido en una pieza muy importante del exocerebro. Además de su función primaria de permitirnos hablar a distancia con otras personas, nos sirve para navegar por Internet, enviar y recibir toda clase de mensajes por WhatsApp, Skype, Twitter o Facebook, escuchar música, leer periódicos y libros, consultar mapas para orientarnos, dirigirnos por el tráfico urbano, comprar muchas cosas, pedir un taxi en Uber, rentar una casa en Airbnb y muchas cosas más. Es un pequeño exocerebro electrónico que conecta nuestro sistema nervioso central con un amplio universo social y cultural. Es una pieza importante del bienestar potencial de millones de personas, aunque también canaliza tensiones y malestares. El aparato es también una especie de robot inteligente que, aunque forma parte de nuestra conciencia, por sí mismo no es consciente.

Podría decirse que el teléfono celular inteligente es una versión hiperdesarrollada de los amuletos y colgijes a los que se refería el médico Qusta ibn Luqa, del que hablé al comienzo de este ensayo. El teléfono móvil concentra una gran cantidad de funciones simbólicas y genera un efecto placebo parecido al que producen, según Qusta, los conjuros y encantamientos. El teléfono celular es un amuleto portátil como los que

producían efectos curativos en tiempos antiguos, dotado de un inmenso poder simbólico. Seguramente cuando algún paciente de Qusta perdía su amuleto colgante sentía una gran ansiedad, parecida a la llamada *nomophobia*, un término acuñado en 2010 cuando se analizaron los resultados de una investigación del servicio postal británico sobre las tensiones que produce la separación o pérdida del teléfono móvil en los usuarios de esta prótesis. El término abrevia la expresión «*no mobile phone phobia*» y se refiere al temor de no estar en contacto con el entorno por haber perdido o estar separado del aparato móvil. En español la *nomophobia* podría llamarse sintelcelosis (sintelcelicia o sintelcelitis). Por supuesto, el teléfono celular es mucho más que un amuleto al que se le atribuyen fuerzas benéficas. Y los efectos nocivos que produce la falta de contacto con la prótesis electrónica van más allá del llamado efecto nocebo. Este efecto nocivo, como he dicho, es lo contrario del efecto placebo: resulta de la creencia de que ciertos medicamentos u operaciones médicas producen resultados dañinos, a pesar de que se trata de sustancias inocuas o de cirugías simuladas. Por supuesto, aquí también hay una similitud con los daños que, por ejemplo, puede ocasionar la creencia de que un brujo puede echar el mal de ojo contra alguien. El hecho es que la manipulación de los sistemas simbólicos a los que están conectadas las personas puede producir efectos benéficos o dañinos en ellas.⁴⁶

Aparentemente nos enfrentamos al problema de los efectos de las sensaciones subjetivas en la operación de la conciencia. Hace años un especialista en el consumo de mercancías observó que las posesiones de las personas son vistas como parte de su identidad. Russell Belk, siguiendo el antecedente de William James, afirmó que los humanos son lo que poseen.⁴⁷ En realidad se refirió principalmente a la sensación subjetiva de la extensión del yo y no se preocupó por investigar si las posesiones de un individuo podían ser *realmente* una extensión de la conciencia, como yo creo que es el caso. En esta misma línea, investigaciones recientes sobre los impactos de los medios de comunicación han llegado a la propuesta de que, así como hay un iPhone, hay también un iSelf, un iEgo. La partícula i, usada por la célebre marca de teléfonos inteligentes, sirve para denotar la extensión de la conciencia por medios electrónicos. Un estudio ha mostrado la influencia fisiológica y psicológica de la falta de acceso al teléfono inteligente (la sintelcelosis). Las pruebas cognitivas mostraron una

calificación más baja y una mayor ansiedad provocadas por la falta de móvil, que también ocasiona un pulso más acelerado y una presión sanguínea más alta. Cuando los participantes en este experimento recuperaron su teléfono, mejoraron sus calificaciones y se normalizaron su pulso y su presión. Se concluye que la extensión de la conciencia mejoró la situación de los sujetos de esta investigación, quienes consideraron el aparato telefónico como una ampliación de su yo. La privación de teléfono celular provocó en los sujetos estudiados resultados pobres en las pruebas cognitivas y mayor ansiedad fisiológica.⁴⁸ Como el teléfono inteligente es un robot que extiende los poderes de nuestro ego, la exploración de estas máquinas será el tema de la segunda parte de este libro, donde indagaré si ellas serán capaces de desarrollar por sí mismas una forma artificial de conciencia.

Es sabido que los mecanismos neuronales del placer son mucho menos comprendidos que los procesos del dolor. Son mucho más estudiados los efectos analgésicos y de alivio que producen los placebos, en contraste con el escaso interés por estudiar sus efectos hedónicos. Unos investigadores que se interesaron por este último enfoque descubrieron que los efectos de alivio son solo la mitad del problema. La otra mitad son los efectos placenteros, la hiperhedonia.⁴⁹ Lo que han encontrado es que los efectos analgésicos y los placenteros activan los mismos circuitos cerebrales. El placer y el dolor muestran similitudes en términos neuroquímicos y neurofisiológicos, concluyen en esta investigación, que se propuso comparar el proceso cerebral de la hiperhedonia con el de la analgesia en los efectos del placebo. El resultado señala que el placer incrementa la activación sensorial mientras que los efectos analgésicos del placebo producen un decremento de la activación sensorial. La hiperhedonia y la analgesia parecen compartir los mismos mecanismos. Ya mencioné más arriba que los efectos analgésicos son mediados por el sistema opioide, así como por las vías que modulan el dolor en el sistema nervioso central.

Los rituales del placer nos llevan a la antigua idea de la catarsis, noción que se asoció en Grecia con el gran espectáculo del teatro. La metáfora de la catarsis tiene su origen en la medicina hipocrática, que recomendaba la eliminación de los malos humores corporales para sanar al enfermo y restablecer el equilibrio en su cuerpo. Pero las purgas aplicadas por los antiguos médicos, mediante sangrías, eméticos y laxantes, no eran un

método placentero de sanación, aunque se esperaba que después del tratamiento debía llegar el alivio. Aristóteles, en cambio, tomó la idea de catarsis para asociarla con el placer que producían en el teatro las representaciones de tragedias. En su *Poética* la catarsis tiene una connotación metafórica importante: se trata de una puesta en escena que produce emociones que purifican a los espectadores mediante un placer que conmociona. La tragedia limpia a los espectadores de los mismos temores y horrores que representa. Es un placer trágico que purifica y sana al público de las aflicciones y los temores que lo afectan. En su *Política* Aristóteles se refiere a los efectos catárticos de la música. A veces la música produce en los que escuchan un frenesí religioso que sana a los que participan en el ritual orgiástico. Otras veces la música simplemente proporciona un placer inocente e inocuo que relaja a la gente.

El concepto de catarsis fue tomado por el médico Josef Breuer para referirse a una práctica curativa que consistía en dejar a los pacientes neuróticos hablar libremente bajo hipnosis, con el objeto de que descargasen sus emociones. Este método fue retomado por Freud, quien lo desarrolló y eliminó la hipnosis en el tratamiento. Así nació el psicoanálisis. Freud reconoció que en realidad se trataba de un método antiguo consistente en estimular en el enfermo los estados psíquicos favorables a la curación. Señaló que los pueblos primitivos usaban casi exclusivamente la psicoterapia para curar a los enfermos. Los brebajes, las fórmulas mágicas, las abluciones purificadoras o la provocación de sueños proféticos, dijo Freud, actuaban terapéuticamente debido a que eran acompañados de un tratamiento psíquico: la magia de la palabra, que podía eliminar manifestaciones morbosas, especialmente aquellas basadas en estados anímicos. Así comenzó lo que Freud llamó el «tratamiento por el espíritu» (*Seelenbehandlung*).⁵⁰ Esta acción terapéutica se basaba en el uso de la palabra, mediante la sugestión, y su origen fue la técnica de Breuer, la «sugestión hipnótica». Freud al respecto recuerda un dicho antiguo según el cual las neurosis no son curadas por la medicina sino por el médico, cuya personalidad ejerce un fuerte influjo psíquico. Alude a la sugestión hipnótica y a las terapias por distracción y entretenimiento, pero señala que se inclina por el método que Breuer denominó *catártico* y que él prefiere llamar *analítico*.⁵¹ Freud no usó el concepto de placebo para referirse a la magia de la palabra o a la catarsis. Hoy muchos creen que el psicoanálisis

es en realidad una manifestación del efecto placebo. Pero en el tratamiento analítico quien principalmente toma la palabra es el enfermo, no el médico. Hay que señalar, no obstante, que es el médico quien encabeza el ritual en un espacio, el consultorio, especialmente preparado para estimular la catarsis. Lo importante es que el paciente crea firmemente en las virtudes y la eficacia del proceso catártico dirigido por el analista, aunque sea difícil y doloroso.

La psicoterapia analítica parte de la idea, un tanto dudosa, de que el ritual terapéutico permite que lo inconsciente fluya hacia la esfera de la conciencia. Freud escribe que «la emergencia de lo inconsciente va enlazada a sensaciones de displacer, a causa de las cuales es rechazada». Por ello el médico, continúa Freud, debe «llevar al enfermo a aceptar algo que hasta entonces había rechazado (reprimido) a consecuencia de la regulación automática determinada por el displacer».⁵² El disgusto que provoca la catarsis psicoanalítica se parece más a las desagradables purgas hipocráticas que al placer trágico aristotélico. Es posible que, en realidad, durante el ritual psicoanalítico haya un flujo de lo consciente hacia la esfera de lo inconsciente, pues es precisamente el convencimiento casi mágico de que el tratamiento es benéfico lo que estimula los procesos neuroquímicos (obviamente inconscientes) que provocan el bienestar del paciente.

Hay que señalar que en el ejemplo que expuse más arriba, el exitoso tratamiento de una esquizofrénica realizado por la psicoanalista Marguerite Sechehaye, hay un cambio notable en el método. El análisis catártico encaminado a hacer hablar a la enferma solamente produjo un alivio pasajero, y fue preciso alterar el tratamiento con una nueva técnica: la realización simbólica de los deseos inconscientes de la esquizofrénica. Fue necesaria la activa intervención hablada, ritual y simbólica de la psicoanalista para lograr la sanación de la enferma.

El efecto placebo debe operar sobre la base de la conciencia, no del inconsciente. En este sentido, aunque podemos sospechar que en el psicoanálisis está muy presente el efecto placebo, sus presupuestos teóricos van en un sentido muy diferente, pues parten del postulado de que durante el tratamiento hay un flujo verbal procedente de memorias reprimidas almacenadas en el misterioso reino del inconsciente. No me ha interesado aquí hacer una crítica de las teorías freudianas. Lo que me importa es haber mostrado que el efecto placebo es una prueba de que las estructuras

simbólicas alojadas en las redes sociales y culturales no solamente ejercen una influencia en el funcionamiento del cerebro, sino que son además una parte constitutiva de la conciencia de los humanos, una conciencia montada tanto en el tejido neuronal como en las texturas simbólicas que nos rodean. La reflexión sobre los muy extendidos rituales del placer y la palabra es una forma de enfrentar y explicar el misterio de la conciencia humana.

7. ZOMBIS Y TRANSHUMANISTAS

Quienes tratan de entender el fenómeno de la conciencia le dan una gran importancia a la sensibilidad, sea que esta se manifieste como placer, dolor o de otras muchas maneras. Se ha insistido en que el gran problema de la relación entre la mente y el cuerpo se puede entender mejor como la relación entre las sensaciones y las funciones. De aquí se desprende una pregunta que les parece fundamental a muchos filósofos y para la cual no se ha encontrado respuesta: ¿por qué en lugar de tener sensaciones no realizamos solamente funciones? ¿Por qué el dolor o el placer en lugar de ser sentidos no son únicamente «funcionados»? ¿Por qué no somos como zombis, esos seres insensibles inventados por los filósofos que solo perciben señales que activan respuestas funcionales? Si no contestamos estas preguntas no lograremos, se cree, descifrar el misterio de la conciencia. Es necesario saber cómo y por qué tenemos sensaciones, y no solamente comprobar que hay correlaciones entre una determinada conducta y ciertas sensaciones.⁵³ Les parece a muchos que, por ejemplo, cuando traducimos una sensación de dolor para comunicarla mediante el habla y avisar a otras personas que estamos seguros de que pueden tener la misma sensación desagradable, no estamos entendiendo el cómo y el porqué de la conciencia, debido a que nuestra descripción simbólica del dolor podría funcionar igualmente si nuestras quejas no correspondiesen a ninguna sensación. Solo nosotros sabemos que sentimos dolor, pues lo sufrimos, pero no sabemos por qué lo sentimos. No sabemos por qué una función cerebral debe sentirse en lugar de solamente «funcionarse».

El argumento del zombi, muy popular en la filosofía anglosajona, ha sido usado para apoyar la idea de que podrían existir unos seres (o máquinas) que actuarían exactamente como nosotros salvo que carecerían de conciencia. Se trata de unos zombis que teóricamente son idénticos a los humanos tanto en su conducta como en su funcionamiento excepto que carecen de experiencias. Hay quienes concluyen que la posibilidad lógica de un zombi insensible e inconsciente demostraría que en los humanos reales hay una sustancia metafísica, como el alma, que alienta a la

conciencia y genera sensaciones. El experimento mental del zombi pretende mostrar que es posible que la conciencia y el funcionamiento del cuerpo tengan una existencia lógica separada.

La fascinación por el zombi parece impulsarnos a formular una pregunta dirigida a Dios. ¿Por qué no somos zombis? ¿Por qué debemos sufrir? Podemos encontrar la respuesta en la Biblia: el sufrimiento es una consecuencia del pecado original. Los pensadores que se obsesionan por la teoría del zombi no se han percatado de que podemos encontrar en la Biblia un antiguo ejemplo de unos seres idénticos a los humanos pero carentes de sensibilidad y conciencia. Me refiero a Adán y Eva antes de que probaran los frutos del árbol prohibido en el jardín del Edén. En mi interpretación, un tanto juguetona, antes de probar el fruto del árbol de la ciencia, Adán y Eva eran seres inmortales y carentes de una conciencia que les permitiese distinguir el bien del mal. Vivían desnudos como los animales sin sentir vergüenza. Acaso Eva no era tan insensible, pues tuvo curiosidad y deseó comer del fruto prohibido, impulsada por la serpiente. Pero me gusta imaginar que los dos habitantes del Paraíso fueron unos zombis libres de penas y de placeres antes de su caída en el pecado. En esta línea de interpretación, san Agustín se dio muy bien cuenta de que si Adán no hubiese pecado habría sido capaz de ejecutar una función como la erección del pene «sin libido ni pasión», de tal manera que usaría su órgano erguido como el hombre que con su mano siembra la tierra, para penetrar en una Eva igualmente insensible y frígida pero funcionalmente capaz de procrear sin dolor.⁵⁴ ¿Por qué los humanos no nos mantuvimos como Adán y Eva, los zombis originales? La respuesta religiosa es evidente: por el pecado que trajo las emociones desconocidas, el dolor, el sufrimiento, la conciencia y, podemos suponer, otras sensaciones como el placer y la alegría.

El argumento del zombi, pues, puede tener un trasfondo religioso. El misterio de la conciencia, así visto, pareciera no tener una solución científica. Pero si consideramos que la conciencia se extiende fuera del cerebro, es probable que nos acerquemos a una explicación. Sin embargo, hay mucha resistencia a esta idea: ¿para qué quiero yo buscar extensiones de mi dolor de cabeza en un lugar fuera del cráneo? Pareciera que el hecho de que puedo nombrar y explicar mi dolor no tiene nada que ver con la sensación subjetiva. Pero los humanos usamos la palabra para referirnos al sufrimiento y a los posibles remedios. El problema no es que los otros solo

saben de mi dolor porque yo se lo explico. El uso del habla es un elemento fundamental de la conciencia. Un animal carente de lenguaje no puede *saber* que siente un dolor: solamente lo sufre, pero no es consciente de ello. Muchos de los otros animales sin duda sufren, pero no saben que sienten dolor. Una cosa es el sufrimiento que va acompañado de palabras y otra cosa muy diferente es el dolor provocado solamente por señales. La confusión proviene de que para neurocientíficos como Stevan Harnad la conciencia es igual a la sensibilidad, mientras que yo sostengo que a las sensaciones se agrega el uso de la palabra y, más ampliamente, de un complejo sistema simbólico. Otros neurocientíficos como Gerald Edelman comprenden que el argumento del zombi es una imposibilidad lógica y que no puede funcionar el proceso nervioso que forma la base biológica de la conciencia sin que al mismo tiempo exista la conciencia.⁵⁵ Así, exigir que se conteste por qué y cómo no somos tan insensibles como un zombi es una forma de invocar un misterio irresoluble.

La conciencia humana sin duda está basada en la sensibilidad propia de un ser vivo. Pero hay una peculiar sensibilidad que es necesario entender y que a su vez puede explicar las sensaciones subjetivas. Cuando los circuitos cerebrales se conectan con redes simbólicas del entorno cultural, al mismo tiempo se percatan del carácter extraño y externo de las formas de comunicación lingüística, visual, musical e instrumental que son características de la sociedad. Esta sensibilidad, que tiene por base la incompletitud neuronal y el reconocimiento de la extrañeza de los símbolos que nos enlazan con otros seres humanos, es la causa de que cristalice un yo individual, único y consciente.

El efecto placebo es una muestra de la forma en que se enlazan la sensibilidad y la palabra. El habla constituye, como lo he mostrado, una poderosa prótesis que en el contexto de rituales médicos o chamánicos puede aliviar algunos sufrimientos y proporcionar placeres y satisfacciones. La idea de que las prótesis rituales y simbólicas pueden crear extraordinarios efectos biológicos y contribuir a la sanación se ha extendido a lo que podría llamarse, si queremos verlo con un poco de ironía, un chamanismo posmoderno: el transhumanismo. Aquí los rituales y los simulacros son sustituidos por prótesis realmente implantadas en el cuerpo de los humanos para aumentar sus capacidades. Las palabras se sustituyen por artefactos. Se trata de lograr una hibridación de lo humano con la

máquina para estimular un progreso cognitivo y biológico mediante la creación de una nueva especie poshumana: los cíborgs. Los médicos podrán así tratar las enfermedades físicas y los desequilibrios psíquicos de manera directa, por medio de un cerebro mejorado y aumentado por la tecnología. El gran impulsor del transhumanismo Ray Kurzweil está convencido de que mediante implantes intracerebrales los humanos podrán convertirse en seres mucho más inteligentes y sanos.⁵⁶ Esta propuesta, que reconoce implícitamente la estrecha relación entre el cerebro y el exocerebro, quiere dar un paso más para incorporar los mecanismos exocerebrales al cuerpo mismo del cerebro mediante implantes tecnológicos hipersofisticados. Pero en el transhumanismo la relación entre el cerebro y el entorno simbólico se convierte en un estrecho vínculo que une el cuerpo humano con artefactos robóticos y cibernéticos que operan con programas. Según Kurzweil no está lejos una condición humana en la que miles de nanorrobots circularán por nuestra sangre y en nuestro cerebro para combatir agentes patógenos, corregir errores en el ADN, eliminar toxinas y realizar tareas que mejorarán nuestro bienestar físico. Habrá una interacción entre nanorrobots y neuronas que deberá sustituir estas últimas por versiones no biológicas más duraderas y más eficientes.

Los transhumanistas suelen pensar que la singularidad tecnológica que producirá unos seres poshumanos no está muy lejana. Creen que dentro de apenas unos decenios estaremos presenciando estas formidables mutaciones. Confían en la aceleración exponencial de las tecnologías biocibernéticas. Creo que se equivocan en cuanto a la proximidad de las mutaciones; de momento los transhumanistas parecen más bien unos chamanes que viajan al futuro y predicen la sustitución de órganos por prótesis tecnológicamente sofisticadas con el objeto de llegar a una condición utópica. Hay un ingrediente religioso en la espera del advenimiento de la Singularidad, con mayúscula, que abrirá la puerta a una nueva época.

Lo que nos hace humanos es otra singularidad, con minúsculas, la de las prótesis artificiales que constituyen la cultura y el entorno social que los humanos hemos creado. La singularidad que reúne en una sola red la palabra con la sensibilidad. La nueva Singularidad de los transhumanistas parece consistir en la transformación de porciones de nuestro exocerebro artificial en una nueva artificialidad mecánica interna, implantada con el

objeto de «mejorar» y «aumentar» las capacidades humanas. En su forma extrema, esta transformación sería una reducción sintética de partes del espacio social y cultural simbólico que nos rodea a un conjunto de dispositivos cibernéticos *internos*. El «hombre aumentado» del transhumanismo estaría dotado de una continuación de la artificialidad propia del exocerebro por medios electromecánicos y cibernéticos. Los primitivos amuletos y conjuros, junto con los rituales médicos que aumentan las potencialidades curativas de los fármacos y las cirugías, serían el embrión que daría paso a los nuevos humanos del futuro, quienes gracias a la magia de la tecnología interiorizarán las propiedades del exocerebro, marginarán el cuerpo biológico y eliminarán las formas de conciencia individual que conocemos. Posiblemente serían unos nuevos zombis, desposeídos de la molesta sensibilidad que nos impone el cuerpo biológico. Para que esta nueva condición poshumana ocurra deberá emerger una forma de conciencia artificial, un problema que abordaré en la segunda parte de este libro.

II. La construcción de una conciencia artificial
Antropología del efecto robótico

1. EL MISTERIO DE LAS MÁQUINAS PENSANTES

Los circuitos exocerebrales de la conciencia humana son totalmente artificiales. No tienen un carácter orgánico. Forman parte de la cultura y de las extensas redes sociales que nos rodean. Pero son capaces de producir efectos en la operación de los circuitos neuronales, como he comprobado al reflexionar sobre la influencia de los placebos en el cerebro. Los artificios del ritual médico o chamánico y de sus simulacros generan efectos en la bioquímica neuronal. Algunos segmentos de los circuitos exocerebrales son prótesis mecánicas dotadas de inteligencia artificial, como las computadoras o los teléfonos móviles. Cuando prótesis electrónicas y digitales inteligentes se conectan al cerebro, se transforman en una parte artificial e inorgánica de nuestra conciencia. Es evidente que el estudio de la manera en que se comportan estos segmentos artificiales de la conciencia tiene una gran importancia en la comprensión de las posibilidades de expansión de máquinas dotadas de una poderosa inteligencia artificial capaces de alcanzar niveles humanos e incluso superhumanos. Aquí surge la pregunta: ¿las máquinas caracterizadas por tener una potente inteligencia podrán algún día desarrollar una conciencia y una autonomía similares a las humanas? Quiero emprender un vuelo imaginario al futuro para explorar las posibles contestaciones a este interrogante.

Para levantar el vuelo es necesario partir de una teoría sobre la conciencia. Hay quienes, como Susan Blackmore, creen que la conciencia es una ilusión, un fenómeno que existe pero que no es lo que parece ser.⁵⁷ Ella llega a la conclusión de que cualquier máquina capaz de imitación podría adquirir esta clase de ilusión. Se trata, dice Blackmore, de una ilusión creada por memes con el objeto de reproducirse, pero no es otra cosa que la falsa idea de un yo consciente que persiste a lo largo de una vida. Ella ha adoptado la teoría de los memes propuesta por Richard Dawkins, que concibe la existencia de conglomerados de información cultural que funcionan como los genes biológicos.

La teoría de los memes es la base para creer que unos robots que imiten a

los humanos podrían adquirir la ilusión de un yo y de una conciencia, de la misma manera en que lo hacen los humanos. Las máquinas capaces de imitar y copiar podrían albergar un proceso evolutivo de sus memes que las llevaría a convertirse eventualmente en entidades autorreplicantes, sin que este proceso de emergencia sea controlado por los humanos. Tales robots inventarían espontáneamente los procesos de autorreferencia. Se trataría de máquinas que, como nosotros, tendrían la peculiar ilusión de ser conscientes. Una ilusión es algo que engaña o provoca una falsa interpretación. Las razones por las cuales Blackmore cree saber que la conciencia es una falsa interpretación, una ilusión, son oscuras. Cuando le expuso su idea al filósofo John Searle, este la rechazó diciendo que si alguien tiene la ilusión de ser consciente, entonces es consciente. La distinción entre apariencia y realidad no puede aplicarse a la conciencia.⁵⁸ De la definición de la conciencia humana como una ilusión se desprende que los memes que operan en los robots podrían crear la misma falsa impresión de que albergan un yo dotado de continuidad. La idea de que nuestra conciencia es una ilusión nos lleva a tratar de explicar por qué somos conscientes de que nuestra conciencia es una ilusión. Esto nos introduce en una inútil regresión infinita, como ha explicado el filósofo Peter Farleigh.⁵⁹

El problema de esta teoría es que se basa en la operación de unas supuestas unidades básicas de información –los memes– cuya existencia jamás ha sido probada y cuya definición precisa brilla por su ausencia. Pero como la inteligencia artificial y los robots funcionan mediante procesos de copiado, réplica y cómputo de información, parece fácil suponer que algún día aparecerá la conciencia en unas máquinas capaces de imitar y copiar, de la misma forma en que supuestamente los memes hicieron brotar la conciencia de los humanos. Desde esta perspectiva, será cuestión de esperar a que los constructores de inteligencias artificiales, apoyados en la extraordinaria velocidad con que se desarrolla la tecnología y en la cada vez mejor calidad de las máquinas y los programas, lleguen a crear robots tan potentes que serán capaces de evolucionar hacia formas de conciencia similares a las humanas. Es una esperanza legítima que nos lleva a especular si transcurrirán decenios, siglos o milenios para que emerja esa extraordinaria singularidad.

En el polo opuesto tenemos la opinión del matemático Roger Penrose,

quien está convencido de que la conciencia se basa en procesos no computacionales. En consecuencia, para Penrose es imposible que una inteligencia artificial desarrolle una mente consciente como la humana; está convencido de que un algoritmo, mediante una computadora, no puede producir una mente consciente.⁶⁰ John Searle también considera que las máquinas digitales con sus algoritmos son incapaces de comprender lo que hacen. Aunque acepta la idea de que el cerebro funciona como el algoritmo de una computadora, cree que la conciencia solo puede brotar si el programa encarna en un cuerpo biológico y no en una máquina. No quedan claras las razones por las cuales cree que se requiere forzosamente de un hardware orgánico para alojar el software inscrito en el sistema nervioso central.

Estas discusiones muestran, a mi parecer, que la exploración del territorio teórico de la conciencia y la inteligencia artificiales es una buena oportunidad para poner a prueba las teorías que tratan de explicar la manera en que la mente humana crea la identidad personal y dibuja el contorno de ese yo que percibe y siente. Por su lado, también los ingenieros dedicados a construir y programar máquinas inteligentes, si quieren crear conciencias artificiales, necesitan algunas guías teóricas que los orienten en su trabajo. Hay que admitir que la construcción de máquinas inteligentes y conscientes no está obligada a seguir exclusivamente el modelo humano. Pero sin duda el ejemplo de la conciencia humana encarnada en un organismo biológico puede estimular a los nuevos ingenieros de almas.

Desde mi punto de vista, la conciencia no es únicamente un fenómeno biológico. La conciencia es un híbrido que enlaza circuitos neuronales con redes socioculturales. Es una confluencia de señales aparentemente electroquímicas en el cerebro con símbolos culturales en el entorno social. Esta congregación es particularmente importante en los momentos en que los humanos comprueban, no sin angustia, que para sobrevivir o superar dificultades no pueden confiar ciegamente en sus recursos biológicos; tienen que acudir a apoyos extrasomáticos de carácter cultural. En estas circunstancias los circuitos cerebrales se conectan a prótesis culturales. Y en esta conexión saltan las chispas de la autoconciencia, ya que las redes neuronales se percatan de la exterioridad y extrañeza de esos canales simbólicos y lingüísticos. Los circuitos neuronales son sensibles a su carácter incompleto. Me gusta imaginar que tienen una sensibilidad como la

del cangrejo ermitaño, que cuando crece tiene que abandonar desnudo un exoesqueleto ajeno para alojarse en otro más grande. La conciencia humana contiene esta clase de sensibilidad ante la incompletitud de los circuitos cerebrales y al mismo tiempo reacciona ante el hecho de que las prótesis culturales a las que se conecta tienen un carácter extraño, no biológico. El resultado es esa peculiar subjetividad que caracteriza a los humanos y que es tan difícil de explicar. Los resortes cerebrales de esta singular sensibilidad siguen siendo un misterio que los neurocientíficos deben aclarar. Son los resortes de una conexión entre señales biológicas y símbolos culturales, dos esferas muy diferentes que interactúan con precisión en forma asombrosa. No se ha logrado todavía construir una teoría unificada que abarque las dos esferas, de la misma manera que los físicos todavía no logran unificar la mecánica gravitacional con la cuántica.

El tema de la sensibilidad de los robots me recuerda un ensayo del gran escritor Heinrich von Kleist, escrito en 1810, que quiero comentar como una clave literaria para entrar al problema de la conciencia artificial. El dramaturgo alemán vivió obsesionado por el conflicto entre las emociones y la razón. Cuando estudió matemáticas y leyó a Kant perdió la fe en el conocimiento y creyó que la separación entre el yo y su entorno revelaba una trágica división entre el sujeto y el saber: con ello los humanos quedaban condenados a la incertidumbre.

En *Über das Marionettentheater*, un ensayo en forma de diálogo, Kleist cuenta la historia de un muy popular bailarín, amigo suyo, que solía asistir con frecuencia al teatro de marionetas que se presentaba con gran éxito en la plaza del mercado. Los títeres solían danzar con tan extraordinaria gracia que el bailarín consideraba que superaban a los humanos. El titiritero sería como el programador de un robot que con habilidad produce el movimiento de un objeto inanimado. Kleist le comenta al bailarín que el operador de las marionetas puede hacerlas bailar sin tener ningún sentimiento o sensación, como un organillero que da vueltas a la manija de su aparato. El bailarín le replica que no es así, pues los dedos del titiritero se relacionan con los títeres como los números con respecto a los logaritmos o la asíntota con respecto a la hipérbola, aunque creía que los rastros de voluntad humana podían ser eliminados en las marionetas para que su danza fuese generada exclusivamente por fuerzas mecánicas. Algo así ocurre con el ingeniero que

programa un robot: proyecta sus sentimientos en algoritmos que animan los movimientos totalmente inconscientes de la máquina.

El bailarín cree que si un artesano construyese una marioneta con las instrucciones que él ha imaginado, el títere podría bailar mucho mejor que los grandes danzarines humanos. Y, como si hubiese previsto las sofisticadas prótesis robóticas que hoy conocemos, puso el ejemplo de unas piernas artificiales fabricadas por artesanos ingleses para sustituir las que algunas personas desafortunadas han perdido. Quienes las usan son capaces de bailar con una elegancia sorprendente, dice el bailarín, lo que me recuerda a la modelo Aimee Mullins, a quien de niña tuvieron que amputarle ambas piernas, famosa por sus logros atléticos con el apoyo de muy diversas y sofisticadas prótesis. Los títeres que el bailarín tiene diseñados en su cabeza ejecutarían sus pasos con un alto nivel de flexibilidad, ligereza y armonía. Pero la gran ventaja de los títeres es que pueden actuar sin afectación, pues el operador que los controla con sus hilos hace que carezcan de vida, como puros péndulos gobernados solo por la ley de la gravedad, lo que es una excelente cualidad de la que carecen la mayoría de los humanos que danzan. Además, estos títeres tienen la ventaja de ser casi ingravidos y por ello estar libres del mayor obstáculo en el baile: el peso del cuerpo. Igualmente, el ingeniero que construye un robot sabe que la máquina carece de vida y está libre del peso de los males que afectan a los seres humanos.

Los bailarines son demasiado conscientes de cada uno de sus movimientos. Kleist le contesta que nunca creará que un títere mecánico pueda tener más gracia que un cuerpo humano vivo. El bailarín le contesta que, en materia de gracia y elegancia, un humano jamás puede alcanzar el nivel de una marioneta. La explicación la encuentra en el tercer capítulo del libro del Génesis: la conciencia estorba a la gracia natural, esa inocencia que existía en el Paraíso original, antes de que Adán y Eva comiesen los frutos de árbol de la ciencia. Hay que entender que el espíritu no puede errar allí donde no hay espíritu, como en las marionetas. Así, continuando con el paralelismo, unos zombis inconscientes, unos androides artificiales, podrían actuar mejor y ser más inteligentes que los humanos que los han construido. En el mundo orgánico, dice el bailarín, conforme la reflexión se oscurece y se debilita, la gracia emerge con más decisión y brillo. La gracia brota con gran pureza allí donde no hay conciencia o bien donde la

conciencia es infinita. Es decir, aparece en el títere o en el dios. Kleist le pregunta asombrado si debemos volver a comer del árbol de la ciencia para regresar al estado original de inocencia. El bailarín le dice que sí, pero que ese sería el último capítulo en la historia del mundo. Al año siguiente de escribir este ensayo, Kleist llegó al último capítulo en la historia de su vida, después de sellar uno de los más famosos pactos suicidas en la historia de la literatura. Su amada amiga Henriette Vogel, enferma terminal, aparentemente le pidió que la matara. Se retiraron a la orilla del Kleiner Wannsee, un lago a las afueras de Berlín; allí Kleist la mató de un disparo e inmediatamente después se suicidó.

¿Kleist decidió matarse porque se dio cuenta de que nunca podría ser ni como las marionetas ni como los dioses? Acaso no soportó el peso de su conciencia solitaria y limitada. Por eso el titiritero que describió en su ensayo procura extinguir las chispas de la conciencia que podrían encender al muñeco inanimado. Por el contrario, en los humanos las chispas de la autoconciencia saltan cuando sus circuitos neuronales se conectan con las prótesis exocerebrales. Las reflexiones de Kleist tienen más de dos siglos de antigüedad y sin embargo si las aplicamos a la relación entre humanos y robots adquieren actualidad. Nos hacen meditar, por ejemplo, en un problema ético: ¿es insensato que los humanos quieran construir máquinas dotadas de una conciencia artificial? ¿No es mejor dejar que las máquinas sean inconscientes, como los títeres de Kleist? Él vería con alarma la posibilidad de transmitir a las máquinas inteligentes el pecado de la conciencia y seguramente preferiría que se mantuviesen como robots insensibles.

2. EL EFECTO ROBÓTICO

Cuando un médico o un chamán realiza un ritual simbólico se suele producir un efecto biológico comprobable, el llamado efecto placebo. De manera similar, cuando un ingeniero crea un programa y lo inserta en una máquina produce un efecto mecánico útil. Lo podríamos llamar el efecto robótico. En ambos casos hay un artificio intelectual que provoca efectos tangibles en un cuerpo, sea orgánico o mecánico. En el caso del efecto placebo hay un requisito indispensable sin el cual no opera: el sujeto cuyo cuerpo se quiere sanar debe ser *plenamente consciente* de que está siendo sometido a un proceso de curación. Pero la persona *no es consciente* de que se le están administrando fármacos inocuos. En contraste, la máquina dotada de un programa de cómputo *no es consciente* de lo que está sucediendo, lo cual no impide que se produzca un efecto robótico y el artefacto se ponga a funcionar encaminado a un objetivo predeterminado. La comunidad científica no cree que los robots, tal como los conocemos hoy, tengan una conciencia similar a la humana (y ni siquiera parecida a la de un ratón o un conejo).

Los robots que conocemos son dispositivos que pueden realizar tareas de forma relativamente autónoma, perciben su entorno y logran actuar intencionalmente porque tienen incorporada una inteligencia artificial creada por el ingenio humano.⁶¹ Los robots no suelen ser capaces de obtener en forma autónoma la energía que necesitan para sustentarse, no se autorreparan ni crecen físicamente y tampoco se reproducen fabricando copias de sí mismos. No es imposible que puedan lograr estas proezas y de hecho ya hay casos en los cuales estas incapacidades se intentan superar. Por ejemplo, existe un robot, llamado EcoBot III, que tiene un sistema digestivo artificial que funciona con una celdilla de combustible microbiano capaz de generar energía eléctrica a partir de una reacción bioquímica. Este sistema digestivo se alimenta de moscas. En su primera versión, este EcoBot, capaz de moverse con la energía eléctrica autogenerada, debía ser alimentado manualmente. Pero en la última versión el aparato tiene un sistema sofisticado para atraer a las moscas mediante colores y olores. Las

moscas atrapadas caen en un depósito que contiene un caldo bacteriano capaz de digerir y degradar al insecto para generar la electricidad necesaria para alimentar su motor. El EcoBot se desplaza con ruedas sobre un charco de agua en un tanque cerrado lleno de moscas. Como necesita agua para su metabolismo, la absorbe mediante un tubo. Este modelo experimental es el único robot existente que es capaz de conseguir en forma autónoma la comida que necesita para generar energía suficiente para moverse y trabajar.⁶²

La autorreparación física de un sistema robótico es algo que de momento solo existe en la imaginación. Los robots no se curan a sí mismos ni reparan una pieza si se les estropea. Tampoco crecen físicamente con el tiempo. Sin embargo, a nivel puramente informático, sin conexión a máquinas operativas, es posible crear programas de crecimiento y autocuración. A este nivel es posible también lograr la reproducción mediante copiado y réplica de datos. Se trata de modelos computacionales muy sofisticados y complejos capaces de imitar muy diversos procesos biológicos, neuronales y sociales.

Veamos un ejemplo que es más que un modelo o una simulación, pues tiene una utilidad práctica y precisa. Se trata del llamado Agente de Distribución Inteligente (IDA, por sus siglas en inglés), desarrollado para la Marina de Estados Unidos.⁶³ Al finalizar cada ciclo de tareas, es necesario asignar a los marinos un nuevo conjunto de deberes. Para ello se usa el sistema IDA que automatiza totalmente la distribución y asignación de tareas. Para ello IDA se comunica por correo electrónico con los marinos e inicia incluso conversaciones normales con ellos mediante cartas que parecen escritas por humanos. Como agente, IDA debe adecuarse a las necesidades de la Marina, a las reglas establecidas, a los costos estimados y a los deseos o necesidades de cada marino. IDA es capaz de procesar mensajes que le llegan en lenguaje natural. A su manera, entiende, reconoce y clasifica el contenido de los mensajes. Su «comprensión» es muy elemental y superficial, pero es suficiente para procesar la información. IDA tiene memorias y un espacio de trabajo donde procesa la información que recibe. Incluso es capaz de simular emociones, como cuando muestra ansiedad por no convencer a un marino de que acepte una tarea, o de manifestar frustración cuando no entiende un mensaje. El sistema opera con muchos miniprocesos relativamente independientes y con propósitos

especiales (llamados *codelets*). Gracias a ello IDA puede crear posibles escenarios y planes de acción para después elegir entre ellos el más adecuado. Así, IDA le ofrece al marino una opción y es capaz de negociar cambios para satisfacerlo. Con el trabajo que IDA realiza va aprendiendo mediante procesos asociativos.

El creador de IDA, Stan Franklin, no le atribuye a su artefacto una conciencia fenoménica. Pero cree que la atribución a las máquinas de una conciencia de este tipo es meramente una actitud de los que observan el artefacto. Se trata solamente de un problema de atribución, por lo que predice que habrá algún día agentes programados suficientemente inteligentes, capaces y comunicativos como para que la gente asuma simplemente que son artefactos conscientes. Entonces el problema de la conciencia de una máquina ya no será relevante.⁶⁴

Otro experto en robótica, Alan Winfield, concluye de manera muy similar sus reflexiones. Está convencido de que, en principio, será posible en el futuro construir un androide tan perfecto como Data, el famoso robot de la serie *Star Trek: The Next Generation*. Estos robots humanoides podrían ser diseñados para comportarse como si fueran conscientes de manera tan convincente que deberían ser contemplados como si realmente fueran conscientes. Pero Winfield agrega una nota irónica: una imitación tan perfecta sería capaz de discutir con un filósofo y convencerlo de que no es verdaderamente consciente y que es solamente un robot. Una máquina como esta se ganaría nuestra confianza, nuestro respeto y un reconocimiento a su identidad personal. Y sería un reto para nuestras profundamente enraizadas ideas sobre lo que significa ser humano.⁶⁵ Desde la perspectiva de la ingeniería robótica no es necesario esperar a que los filósofos o los neurocientíficos resuelvan el misterio de la conciencia. De hecho, puede ocurrir que los ingenieros contribuyan a resolver este misterio al construir robots cada vez más hábiles e inteligentes.

Desde una perspectiva tecnológica, la construcción de robots no requiere de precisión terminológica. Vincenzo Tagliasco ha escrito que, a diferencia del trabajo científico (que requiere definiciones), la tecnología supera las ambigüedades cuando los diseños aterrizan en la construcción de prototipos que tienen que funcionar correctamente en el mundo real. Para construir una máquina con una «conciencia mínima» se requiere que las experiencias del artefacto lo empujen a buscar nuevas experiencias, a mantener un

registro de ellas, a unificar la información y a ser capaz de definir nuevas motivaciones que tengan los mismos objetivos predefinidos que fueron diseñados e inscritos materialmente en el aparato. Tagliasco concluye que la conciencia artificial no será fruto de una disciplina científica, sino que surgirá como consecuencia de un trabajo tecnológico más cercano a la robótica que a la neurociencia o la psicología.⁶⁶

El objetivo, desde esta perspectiva, consiste en construir un robot aparentemente consciente que se comporte como tal. La robótica puede hacer a un lado las emociones, los placeres, los sentimientos, los dolores y los deseos, olvidarse de la sangre, la carne y las hormonas, para concentrarse en construir máquinas cada vez más sofisticadas que algún día llegarán, se espera, a generar una extraordinaria singularidad tecnológica: el surgimiento de inteligencias artificiales conscientes mucho más poderosas que el intelecto con que la naturaleza ha dotado a los humanos.⁶⁷ El efecto robótico se genera cuando una máquina controlada por un programa se transforma en un artefacto que muestra una inteligencia y un comportamiento aparentemente humanos. En su versión más avanzada – imaginada, deseada y temida, pero todavía muy lejana– mostraría todos los síntomas de poseer una conciencia muy similar a la nuestra, sin que se pueda asegurar que verdaderamente el robot no sea más que un zombi muy eficiente. En el efecto robótico el ingenio humano construye unas prótesis complejas e inteligentes que forman parte de la conciencia de sus creadores. Los ingenieros sueñan con que estas máquinas logren en el futuro una autonomía que liberaría a las prótesis de sus dueños y creadores. Al adquirir formas de conciencia autónoma, los artefactos esclavizados se liberarían de sus amos.

3. ¿CÓMO SE EDUCA UN ROBOT?

El problema de la autonomía de los aparatos gobernados por sistemas de inteligencia artificial es fundamental. ¿Puede un robot aprender en forma independiente a realizar tareas para las cuales no ha sido programado y sobre las cuales no tiene información en su memoria? En 1997 la supercomputadora Deep Blue derrotó al campeón de ajedrez Garri Kaspárov en un juego. Fue un hecho que causó sensación, pues una máquina inteligente logró por primera vez ganar la partida a un campeón mundial. Pero Deep Blue logró ganar gracias a sus ventajas en capacidad de memoria y de cómputo, no porque fuese capaz de aprender. Igualmente, el programa AlphaGo ha logrado ganar en el juego de go a varios campeones mundiales. Pero los ingenieros y programadores de Google DeepMind crearon un sistema llamado AlphaGo Zero que es capaz de derrotar a las versiones previas del programa. Lo logra gracias a que es capaz de aprender a jugar sin ayuda humana alguna a partir de cero. La máquina carece de una memoria en la que se hayan instalado previamente modelos de partidas. Aprende sola durante juegos en los que se enfrenta a sí misma, hasta que alcanza un altísimo nivel que le permite ganar a los sistemas anteriores que funcionan gracias a que se les han implantado en la memoria millones de movimientos posibles a partir de partidas jugadas por humanos. Otro aspecto del comportamiento de la máquina superinteligente fue mencionado por Kaspárov cuando perdió: al menos Deep Blue no disfrutó su triunfo.

El algoritmo de aprendizaje que usa AlphaGo Zero ya había sido probado con un sencillo juego de Atari, el Breakout, que consiste en rebotar con una raqueta una pelotita para que pegue a los ladrillos de una pared; al ser golpeado, el ladrillo desaparece y se ganan puntos. DeepMind comienza el juego a partir de cero, sin saber nada. Solamente tiene la instrucción de maximizar su puntuación. Se le van introduciendo series de números que son especificaciones de cómo la pantalla se colorea. DeepMind contesta al azar con otros números que son códigos de las teclas en la computadora que hay que presionar. Lo que vemos es cómo una pelotita es lanzada y cómo la raqueta manejada por DeepMind la rebota. El sistema no sabe nada de esto.

Al comienzo juega muy mal, pero paso a paso va aprendiendo la forma de mover la raqueta para devolver el pelotazo contra el muro. Pronto aprende que gana puntos al pegarle a los ladrillos y rápidamente llega a una estrategia brillante que los mismos programadores no habían concebido (cavar un túnel en la esquina izquierda para que la pelota entre y rebote repetidamente contra el muro y lo destruya). La computadora aprende a jugar por sí misma.⁶⁸

Los programadores de DeepMind crearon en 2019 un sistema que es capaz de superar a competidores humanos en el videojuego Q3A (*Quake III Arena*), que implica a «agentes» que operan sobre mapas y ambientes generados al azar y compiten por capturar banderas de los enemigos y liquidarlos. Es un juego colectivo en el que se enfrentan dos equipos. El sistema de inteligencia artificial entrena a un equipo de agentes o bots que se enfrenta a los humanos. Los bots son capaces de aprender a jugar con gran maestría y vencen a los humanos. Estos agentes artificiales se alimentan exclusivamente de píxeles y puntos ganados en competencias entre ellos mismos, durante las cuales aprenden a jugar en entornos complejos, variados y llenos de otros agentes con los que cooperan (los de su equipo) o se enfrentan (los del equipo enemigo). Cada bot ha aprendido de manera diferente a los demás y por lo tanto tiene un estilo propio. El contexto en el que operan los jugadores no es la sencilla cuadrícula de un tablero de ajedrez o de go, sino un ambiente sorpresivo, contingente y variado lleno de obstáculos y de otros contrincantes cuyas reacciones no son fácilmente predecibles. La clave de la superioridad de los jugadores artificiales radica en que han recibido un entrenamiento previo consistente en jugar miles de veces entre ellos, a la manera en que AlphaGo Zero aprende jugando consigo mismo.

Estos nuevos sistemas parten de un proceso denominado aprendizaje profundo por refuerzo (*deep reinforcement learning*). El refuerzo consiste en dar una recompensa si se alcanza un objetivo, lo que impulsa al sistema a lograrlo de nuevo, de la misma manera en que se entrena a un perro a aprender cosas nuevas, dándole una sabrosa galleta si tiene éxito. Este aprendizaje es «profundo» porque contiene varias capas; la primera capa recibe la información (*input*) y la última le da salida (*output*). Las capas intermedias u ocultas van transformando en forma no lineal la información y la transmiten a la siguiente capa.

Las mismas herramientas diseñadas para el aprendizaje profundo han sido usadas para que una máquina sea capaz de reconocer imágenes. Se trata de una red neuronal circunvolucional (*convolutional neural network, CNN*) que, al igual que las máquinas jugadoras, no necesita ser programada para reconocer peculiaridades específicas de las imágenes que se le presentan. Por ejemplo, la red es alimentada con una serie de muchos animales que incluyen perros de dos razas. La red organiza la información que recibe a través de capas, como es usual en el aprendizaje profundo. En las primeras capas superficiales aprende formas simples y contornos en las imágenes. En las capas más profundas se aprenden conceptos abstractos y complejos, como orejas, colas, lenguas y texturas de las pieles. Una vez que se entrena, la máquina puede reconocer fácilmente las diferentes razas. El proceso CNN puede ser también usado en los vehículos autónomos sin conductor para que puedan reconocer bien a peatones. En medicina la herramienta escanea imágenes y las interpreta, lo que es muy útil en lugares donde no hay personal entrenado para hacerlo. Y, como una aplicación de teléfono móvil, puede hacer diagnósticos rápidos y baratos, como reconocer cáncer en la piel.⁶⁹ Por otro lado, Facebook emplea el aprendizaje profundo en la emisión de recomendaciones y avisos, YouTube lo usa para hacer propuestas a los usuarios y Siri, la asistente personal de voz, lo aprovecha para responder preguntas.

En el videojuego Q3A el jugador artificial, que se entrena mediante aprendizaje por refuerzo, se enfrenta al problema de coordinar lo que parece mejor en cada momento con el objetivo final de ganar el juego. Para superar este problema el bot opera en dos capas paralelas en las que obtiene diferentes recompensas. La capa externa es una especie de metajuego en el que se da preferencia al objetivo final de ganar el juego. En la capa interna el sistema busca obtener ventaja en cada momento, pero se orienta por la estrategia general establecida en la capa externa. Los programadores de DeepMind introdujeron un criterio evolucionista, de tal manera que durante el entrenamiento, después de cada ronda, los peores agentes son eliminados y sustituidos por unas «mutaciones». Después de miles de juegos, tenemos el resultado de una selección de los mejores agentes jugadores. Lo novedoso de este sistema, llamado FTW, es que implica a bots capaces de jugar en equipo, colectivamente. Incluso pueden jugar en equipos mixtos donde participan humanos. El nombre del sistema proviene de unas siglas

del argot usado en Internet para designar lo que triunfa (*For the Win*). Pero también quiere decir *Fuck the World*: chingar o joder el mundo. ¿Una broma sobre la amenaza que significan los robots?

Hay una dimensión de la inteligencia artificial donde se han producido avances notables, apoyada en el aprendizaje profundo. Me refiero a los sistemas traductores de lenguas, entre los cuales destaca el de Google. La calidad de las traducciones automáticas ha mejorado mucho. Pero basta jugar un poco con el traductor de Google para llegar a la conclusión de que el sistema no entiende lo que está procesando. Por ejemplo, si partimos de un texto en español y lo traducimos al húngaro, para enseguida traducirlo de esta lengua al chino e inmediatamente al español de nuevo, veremos por los cambios ocurridos que el procesador de lenguaje no ha entendido nada. No es de sorprenderse: ya John Searle había usado su famoso argumento de la habitación china para mostrarlo. Por ejemplo, la frase inicial en español fue la siguiente: «La muchacha que camina por el bosque ama las mariposas.» Después del circuito a través del húngaro y el chino la frase quedó así: «La niña en el bosque parece una mariposa.» Obviamente una parte importante del significado se perdió por el camino debido a que el sistema traductor no entiende lo que hace. Carece de conciencia. Lo mismo sucede con las máquinas que reconocen imágenes o que juegan go. El carácter autónomo o automático no implica que el procesador sea consciente.

La autonomía en el proceso de aprendizaje es un logro muy notable. Aunque las máquinas no adquieren una conciencia similar a la humana, sin duda muestran en ciertas tareas una inteligencia superior a la nuestra, si pensamos que la inteligencia consiste en la habilidad para lograr objetivos complejos. Esta clase de inteligencia está basada en información y computación, pero no incluye las formas emocionales e intelectuales cuyas finalidades no se pueden definir con precisión y que se consideran asociadas a un cuerpo biológico. La inteligencia artificial no parece albergar o ligarse a las emociones, las imaginaciones, los placeres y las fantasías que impulsan a un intelecto humano.

La inteligencia artificial que computa y procesa información se ha expandido extraordinariamente y sigue desarrollando formas cada vez más sofisticadas. Uno de los aspectos de esta inteligencia artificial supera con creces a la humana: el espectacular crecimiento de la memoria de las máquinas. Y la velocidad de cómputo se está también multiplicando en

forma exponencial. Se han creado máquinas especializadas que realizan tareas mucho mejor que los humanos: son más rápidas, más precisas y más eficientes. Pero, como observa Murray Shanahan, hasta ahora se ha fracasado en la invención de máquinas dotadas de una inteligencia polivalente orientada a propósitos generales. Deep Blue solo sabe jugar ajedrez, AlphaGo únicamente puede jugar go y el sistema CNN no puede reconocer más que imágenes. Esta especialización no existe en los humanos, que se aplican cotidianamente a muchas actividades y están dotados de una inteligencia que se adapta a muy diferentes circunstancias.⁷⁰

No sin razón se ha calificado a los sistemas de aprendizaje profundo como el fruto de una importante revolución tecnológica y científica. Se habla de revolución porque el diseño de máquinas capaces de aprendizaje profundo parte de premisas completamente diferentes a las que orientaron, durante la segunda mitad del siglo xx, la construcción de inteligencias artificiales. Este diseño se basaba en la programación secuencial, lógica y simbólica de reglas que guiaban el comportamiento de unos robots que eran incapaces de aprender. Todas sus reacciones estaban prediseñadas en la computadora que los guiaba. En contraste, el nuevo enfoque está basado en algoritmos de aprendizaje y en el uso masivo de información (Big Data) y se inspiran en el funcionamiento de las redes neuronales.⁷¹ Se ha dicho que la inteligencia artificial se orientó tradicionalmente por la llamada «computación simbólica», mientras que las técnicas de aprendizaje profundo se basan en el reconocimiento de patrones en grandes masas de datos. El término «computación simbólica» puede ser confuso, pues en realidad hace referencia a la computación algebraica que manipula variables que carecen de un valor dado y que por ello se manejan como signos. Más adelante me referiré a la confusión que puede surgir de la sustitución de la idea matemática de signos por la de símbolos, que es una noción compleja que tiene un uso muy importante en el estudio de la cultura.⁷²

Las primeras aplicaciones de la inteligencia artificial, cuando se enfrentaron a la incorporación de datos visuales, no lograron avanzar y, por ello, prefirieron la aplicación de teoremas matemáticos en la programación, pues las computadoras digitales son extraordinariamente rápidas en operaciones lógicas que manipulan signos y muy eficaces en la aplicación de reglas. Pero la gran revolución, como señala Sejnowski, consiste en el uso de técnicas inspiradas en el modelo neuronal establecido por Alan

Hodgkin y Andrew Huxley, quienes por ello recibieron el Premio Nobel en 1963, sobre la manera en que las señales en el sistema nervioso son transportadas por impulsos eléctricos que se convierten en señales químicas en las sinapsis. Para quienes impulsan los sistemas de aprendizaje profundo, las sinapsis son los elementos computacionales básicos del cerebro. Como dice Sejnowski, los circuitos neuronales tienen como propósito resolver problemas computacionales.

La esencia del aprendizaje profundo se encuentra en la capacidad de una máquina de clasificar patrones en diferentes categorías, para lo cual es necesaria una gran masa de datos (imágenes, palabras, números, rasgos, etc.) que es pasada por diversos filtros (capas o umbrales).

Por ejemplo, un programa lógico basado en números y signos algebraicos no es capaz de identificar las imágenes de copas y separarlas de otros objetos, pues hay demasiada variedad en las formas que registra. Pero un sistema de aprendizaje profundo sí es capaz de hacerlo, pues reconoce en un gran número de imágenes un patrón (la copa), gracias a que clasifica similitudes y diferencias entre ellas. La clave consiste en la capacidad de reconocer y separar diversos patrones. Otro ejemplo de esta capacidad son los programas que logran separar patrones de voces que provienen de una fuente sonora «ciega», es decir, en la que hay una grabación de un centenar de personas hablando al mismo tiempo en un coctel. Hay un algoritmo que permite separar frecuencias y clasificar las diferentes voces mediante un sistema de filtros.

El proceso de reconocimiento de patrones, modelos, pautas o formas se basa en muy sofisticados métodos estadísticos de clasificación. En el centro mismo del aprendizaje profundo se encuentra el cálculo de probabilidades. Otro ejemplo especialmente interesante es el sistema robótico de diagnóstico de enfermedades que ha aprendido a partir de la acumulación de una masa inmensa de casos clínicos a detectar patrones. Al recibir los datos sobre los síntomas de un paciente, los utiliza para una aproximación estadística a un diagnóstico que resulta tan acertado como el que realiza un buen médico muy experimentado. Pero el robot diagnosticador es incapaz de explicar, como lo haría un médico, por qué ha llegado a una conclusión,

pues ha utilizado complejos procesos de valoración probabilística de cada síntoma sin entender la enfermedad que sufre el paciente.

Los sistemas de aprendizaje profundo usan métodos como la retropropagación de errores (*backprop networks*), el aprendizaje circunvolucional (CNN, *convolutional neural network*) y el refuerzo mediante recompensas (*deep reinforcement learning*). La retropropagación permite que los errores detectados en la capa de salida, al ser contrastados con la realidad, sean retransmitidos a las capas anteriores para corregirlos. Las redes circunvolucionales son un sistema de filtros por los que pasa la información y en el cual la primera capa selecciona el predominio de un rasgo (perfiles, bordes), una segunda capa vuelve a filtrar rasgos, y así sucesivamente. Son capas circunvolucionales con filtros y umbrales que acaban generando una clasificación muy precisa. El refuerzo mediante recompensas es un proceso que genera un lazo (*loop*) con el entorno y premios cuando la decisión de la máquina es acertada. El proceso de enlace se repite a cada paso, pues el sistema estima el posible premio futuro y con ello decide el paso siguiente.

Estos procesos de aprendizaje artificial quieren ser una variante mecánica de la manera en que los humanos aprenden. Quienes los han inspirado tienen la esperanza de que el diseño de estos robots contribuya a entender el funcionamiento del cerebro. El problema radica en que en realidad estamos ante un círculo vicioso, pues estas máquinas imitan unos circuitos neuronales cuyo sistema de señales no ha sido descifrado y suponen que los circuitos computacionales que sí entendemos (pues han sido ingenierados por expertos) son un sistema de información semejante al de nuestro cerebro. Se parte de un postulado (no comprobado) según el cual los sistemas de información con los que operan las conciencias son independientes de su sustrato, biológico o mecánico, en el que se apoyan. Sin embargo, todos los patrones de activación, las correlaciones o regresiones estadísticas, la distribución de probabilidades y los mecanismos de clasificación con los que funciona el aprendizaje profundo no han creado ningún efecto que podamos definir como consciente. Se han logrado simulaciones espectaculares, como es el caso de la robot Sophia, una máquina humanoide creada en 2016 en Hong Kong por Hanson Robotics. Sophia está dotada de algunos algoritmos de reconocimiento facial y de voces, habla y gesticula con cierta naturalidad, su rostro está modelado a partir de la cara de la actriz

Audrey Hepburn y es capaz de sostener conversaciones sencillas, aunque sus respuestas están previamente programadas. Con frecuencia sus contestaciones carecen de sentido. En realidad, es una robot primitiva y no está dotada de nada cercano a una inteligencia general, pero tiene una apariencia muy atractiva. El espectáculo cibernético de Sophia culminó en octubre de 2017 cuando Arabia Saudita le concedió la ciudadanía, con lo que se convirtió en la primera máquina inteligente a la que se le otorgan los derechos propios de una nacionalidad. Aparentemente tiene más derechos que las mujeres en Arabia. Hay quienes consideran que todo sistema integrado de información, como el que anima a Sophia, tiene alguna forma de conciencia, aunque sea rudimentaria.

4. PANPSIQUISMO

Una manera fácil pero dudosa de abordar y aparentemente resolver el problema de la conciencia artificial consiste en postular que en realidad ya hay conciencia hasta en un termostato, que es un aparato elemental que integra información en un solo sistema. El misterio de la conciencia, desde este punto de vista, se resuelve no por declarar que se trata de una ilusión sino por creer que todo sistema integrado de información es, en alguna medida, consciente. El físico Max Tegmark está convencido de que la conciencia es la forma en que la información *siente*, cuando está integrada en un sistema. Adapta su definición a la propuesta de Giulio Tononi sobre los sistemas integrados y unificados de información.⁷³ Esta teoría considera que la conciencia es un sistema informático independiente del sustrato material que la procesa. Lo importante, dice Tegmark, es solo la estructura de la información, no el cuerpo de la máquina donde opera el sistema. Así como hay una sustancia que es capaz de realizar cálculos arbitrarios, el *computronium*, existe también otra sustancia, el *sentronium*, que tiene una experiencia subjetiva: es una sustancia sensible. La conciencia es, según Tegmark, un fenómeno físico que se siente como nofísico porque funciona como las ondas y el cálculo: sus propiedades son independientes del sustrato físico específico en que se encuentra implantada.

Hay en estas ideas una curiosa y nueva expresión del antiguo panpsiquismo, como lo ha admitido el neurocientífico Christof Koch, quien también usa para fundamentar su idea la propuesta de Tononi sobre la información integrada.⁷⁴ Este panpsiquismo abre las puertas a un futuro dominado por la conciencia artificial que, al ser independiente de cualquier sustrato físico, puede desplegarse libremente en todas las dimensiones hasta alcanzar formas extraordinarias y singulares de superinteligencia autoconsciente. Los sistemas integrados de información son proyectados a un futuro donde su singular psiquismo invade todos los ámbitos de la sociedad. Estos sistemas de información autónomos e independientes alcanzarán, se cree, altos niveles de conciencia. No se explica cómo y por qué estos sistemas *sienten*, tienen sensaciones subjetivas. Tegmark señala

que sería un error creer que una inteligencia artificial necesariamente siente como una persona. Se podrían crear formas de inteligencia artificial consciente que jamás tengan experiencias como el hambre, la sed, el miedo o el deseo sexual. Tegmark imagina que los sistemas inteligentes y conscientes programados para tener cualquier objetivo ambicioso seguramente lucharán para conservarse y alcanzar su finalidad. Estos procesos podrían carecer de ese miedo humano a la muerte, rebajar su sentido de individualidad y convertirse en un conjunto conectado de sistemas superinteligentes que se sienta a sí mismo como un solo organismo provisto de una mente en forma de colmena. Estos sistemas integrados *sentirían* que tienen libre albedrío.⁷⁵

Las sensaciones que puede registrar un robot son por ahora una mera especulación que carece de base sólida. Lo único que podemos ver en los robots dotados de inteligencia artificial es que son capaces de percatarse de cambios en su entorno y reaccionar en consecuencia. Eso no es nada nuevo; por ejemplo, un termostato se percata de la temperatura ambiental y, si llega a un umbral predeterminado, reacciona conectando o desconectando un circuito. Se diría que reacciona cuando siente que debe hacerlo pues tiene un objetivo, es sensible y actúa de acuerdo con ello. Podría pensarse que es similar a una espiroqueta o a un paramecio, organismos unicelulares dotados de movimientos y sensibilidad, acaso formados de esa extraña sustancia que Tegmark llama *sentronium*. Pero estamos muy lejos de las formas humanas de conciencia. No es lo mismo una capacidad de percatación o de percepción que la conciencia basada en sentimientos y emociones característica de los humanos.

Otros físicos han planteado algunas formas de panpsiquismo basados en la teoría cuántica. A partir de la famosa paradoja del gato de Erwin Schrödinger, surgieron algunas teorías sobre la conciencia cósmica. El experimento mental de Schrödinger se refiere a un gato encerrado en una caja de acero que está al mismo tiempo vivo y muerto. Según la teoría dominante en 1935, la llamada interpretación de Copenhague, un sistema cuántico se mantiene como una superposición hasta que hay una observación externa que la colapsa en alguno de los estados posibles. Para mostrar lo absurdo de esta interpretación, Schrödinger imaginó a un gato encerrado en una caja cuya existencia está determinada por si hay o no una emisión radiactiva. Debido al principio de indeterminación cuántica, no se

puede saber si está vivo o muerto hasta que se abre la caja para medirla. Así, mientras no haya una observación, el gato en la caja estaría al mismo tiempo vivo y muerto. Habría dos estados contradictorios. Pero si se quiere conocer el estado del gato es necesario abrir la cámara y tomar una medida, lo cual ocasiona que las dos ondas (la del gato muerto y la del vivo) colapsen en una sola, de manera que ahora se sabe cuál es la condición del gato, cosa imposible de determinar antes de la medición. Con esta explicación del colapso, Schrödinger (y con él Einstein) quiso criticar la paradoja cuántica. Años después, el físico Eugene Wigner pretendió dar una solución diferente a la paradoja. Postuló que solamente una persona consciente puede realizar la medición que provoca el colapso de las ondas. Pero no podemos saber si existe este observador vivo (podría estar muerto) midiendo las ondas a menos que haya otro observador consciente vivo que observa al primero, lo que provoca el colapso. Pero el físico Michio Kaku observa que para que la medición ocurra, debe haber un tercer observador que al observar la función ondulatoria determina si el otro observador está vivo o muerto. Aquí nos insertamos en una regresión infinita, pues siempre se necesitará un observador adicional que observe al previo. Como se requiere de un número infinito de observadores para asegurar que todos están vivos, es necesario pensar en una conciencia cósmica, una especie de dios. Wigner concluyó que no es posible formular las leyes de la teoría cuántica de manera consistente sin una referencia a la conciencia cósmica. Desde este punto de vista, las funciones ondulatorias de los colapsos solo pueden ocurrir si la conciencia es una entidad fundamental en el universo. Podemos comprender por qué Wigner hacia el final de su vida se interesó en la filosofía védica de la India. Kaku, contra esta idea de Wigner, observa que sería equivocado creer que, puesto que la conciencia determina la existencia, sería posible que la conciencia, acaso con ayuda de meditación, pueda controlar la existencia. «En la física cuántica», dice Kaku, «la conciencia hace observaciones y por lo tanto determina el estado de la realidad, pero la conciencia no puede escoger anticipadamente qué estado de realidad actualmente existe.»⁷⁶ Otros intentos de resolver la paradoja del gato han postulado que el universo se está fragmentando para dar lugar a un multiverso de realidades paralelas; en unos universos el gato estaría vivo y en otros estaría muerto. Aquí no hay el colapso de las funciones ondulatorias que crean una sola onda, sino una separación constante de

realidades diferentes y simultáneas. Acaso en alguno de estos universos paralelos hay una singularidad que ha provocado el surgimiento de robots autoconscientes. Aquí entramos en el terreno de la imaginación, lo que es inevitable si queremos explorar el territorio de la conciencia artificial. Pero el terreno por el que quiero viajar no se encuentra tan lejos, en algún universo paralelo.

5. UNA CONCIENCIA MECÁNICA

Quiero ahora delinear los requisitos que debería cumplir una conciencia mecánica similar a la nuestra. Para ello es necesario partir de las características de la conciencia humana tal como las he definido en mi teoría sobre el exocerebro y los sistemas simbólicos de sustitución. El primer paso es aceptar que sea posible construir máquinas no especializadas capaces de orientar su inteligencia a propósitos generales. Estamos aquí, esencialmente, ante un problema técnico de acumulación de funciones en un solo sistema. Se requiere dotar a la máquina de una capacidad de aprender a partir de cero. Ante este reto podemos imaginar un desarrollo muy sofisticado de las técnicas de aprendizaje profundo a las que ya me he referido. Ello ahorraría a los ingenieros tener que elaborar programas para cada proceso y le daría autonomía al robot.

Para que una máquina logre evitar las limitaciones de la especialización, dice Murray Shanahan, tiene que ser dotada de sentido común y creatividad. Se define el sentido común como la capacidad de entender la manera en que el mundo opera cotidianamente, y muy en especial el entorno físico y el ambiente social. Es posible que se acerque a ello con la ayuda de los procesos de aprendizaje profundo por refuerzo, que le proporcionen a un aparato inteligente una capacidad de orientarse en el mismo entorno en el que viven los humanos. Desde luego, se trataría de programas de aprendizaje mucho más desarrollados que los que hoy conocemos. En cuanto a la creatividad, entendida como la capacidad de innovar, de generar nuevas conductas, de inventar cosas nuevas o usar cosas viejas de manera novedosa, es también de acuerdo con Shanahan un ingrediente indispensable para que la inteligencia artificial pueda escapar de la especialización y funcionar de manera polivalente, abierta y general. No sin tristeza el propio Shanahan observa que, después de seis décadas, los ingenieros que trabajan en inteligencia artificial han progresado muy poco. Es inquietante pensar que acaso es imposible que la inteligencia artificial alcance niveles humanos.⁷⁷

Evidentemente, en el nivel en que estamos especulando, nos enfrentamos

a problemas meramente técnicos en el proceso de crear mecanismos concretos que puedan ser implementados de manera práctica. Las soluciones técnicas que al final sean inventadas no generarán por sí mismas formas de autoconciencia. Pero es indispensable superar estos problemas que, aunque sean meramente técnicos, han resultado huesos muy duros de roer. Después aparecerán los problemas verdaderamente difíciles de entender: ¿cómo unos aparatos capaces de desarrollar funciones generales podrían adquirir conciencia? Acaso no se ha logrado crear una inteligencia mecánica general y polivalente debido a que quienes están capacitados para trabajar en ello se encuentran mucho más atraídos por desarrollar robots especializados que cumplen funciones muy importantes en la industria, las finanzas, el comercio, el ejército y la administración. Un engendro mecánico abocado a múltiples tareas, un robot milusos, no sería hoy muy útil, pues obviamente en sus primeras versiones sería tosco y sus habilidades estarían muy lejos de las humanas.

La ingeniería robótica abocada a la construcción de inteligencias artificiales suele explorar dos vías diferentes. Por un lado, trata de elaborar mecanismos que imiten las redes neuronales que funcionan en estructuras biológicas. Por otro lado, se avanza desde cero, acumulando soluciones técnicas e integrándolas de acuerdo con principios funcionales completamente ajenos a la operación de un cerebro biológico y de un cuerpo orgánico. Así se avanzó en el diseño de aparatos voladores, se suele decir, cuando se dejó de imitar los movimientos de las aves para diseñar alas fijas con hélices o turbinas impulsadas por motores. Pero también se dice con razón que el cerebro biológico es el único ejemplo conocido que funciona con una inteligencia polivalente general. Desde luego, se puede experimentar con sistemas híbridos que combinen imitaciones de una red neuronal con mecanismos inventados para cumplir una función sin equivalente biológico (el ejemplo más elemental es la rueda, que prácticamente no existe en los organismos vivos).

Aquí quiero introducir una idea que proviene de mi interpretación de la conciencia humana: si se quiere construir una máquina consciente, su sistema debe conectarse a un exocerebro que realice funciones que el aparato no puede realizar por sí mismo mediante algoritmos. Las funciones no propiamente intelectuales que es incapaz de realizar, o solo precariamente, son aquellas que sí puede realizar naturalmente un

organismo biológico: obtener en forma autónoma la energía que requiere, autorrepararse y reproducirse. Estas funciones podrían implementarse con artificios técnicos incorporados a la máquina o bien el robot acudiría automáticamente, más que a un exocerebro, a prótesis exoesqueléticas que le proporcionen lo que le hace falta: energía, reparación y copiado. Aquí seguimos en el nivel de problemas que podrían tener una solución técnica a no muy lejano plazo, pero hoy estamos todavía muy lejos de lograr en los robots un nivel técnico de autonomía como el que he descrito. Estamos aún al nivel primitivo de la pacífica aspiradora robótica cuyos sensores le permiten evitar obstáculos, recorrer al azar su entorno, descubrir que su batería se agota para ir a enchufarse a una estación de recarga y detectar que su depósito de polvo está lleno para avisar a un humano que debe cambiarle la bolsa. O bien estamos al nivel, también primitivo pero sanguinario, de los muy agresivos misiles de largo alcance, capaces de guiarse a sí mismos hacia un objetivo predeterminado sin ayuda humana llevando con gran precisión un explosivo para ser detonado al llegar. Estas máquinas, limpiadoras o destructivas, son artefactos especializados carentes de inteligencia general.

Imaginemos un futuro en el que ya se hayan logrado resolver los problemas técnicos para construir un robot dotado de una inteligencia general y flexible. ¿Cómo podría saltar en esta máquina la chispa de la conciencia? Podría tener lo que algunos han llamado una conciencia funcional y actuar con gran eficiencia creando en su entorno la apariencia o la ilusión de que se trata de una máquina consciente. Pero podemos suponer con seguridad que la misma máquina no se encontraría poseída por la ilusión de ser consciente por el hecho de tener una potente y versátil inteligencia general. Si usamos el ejemplo de los humanos, esta máquina podría ser consciente solamente a partir del momento en que fuese dotada de (o ella misma construyese) un exocerebro: un conjunto de redes simbólicas que, tejidas en su entorno, la envolviesen. Se trataría de una cultura robótica que teñiría todas las relaciones de la máquina con otras máquinas y con la sociedad humana.

6. LA CULTURA ROBÓTICA

El núcleo fundamental de una cultura robótica se hallaría en el proceso que conectaría a una máquina inteligente con su entorno cibernético y humano con el objeto de completar circuitos que ella no es capaz de cerrar. Esto implica suponer que esta máquina se percataría de su incompletitud. Este percatarse de sus carencias y limitaciones debería generar la delimitación de una unidad funcional de las peculiaridades de una máquina que, a pesar de su carácter incompleto, define así una identidad y un conjunto articulado de rasgos e insuficiencias. Los circuitos que no pueden completarse son, en primera instancia, las formas de comunicación con el entorno para lograr el apoyo y la colaboración de otras máquinas o humanos para llevar a cabo determinadas tareas. Esta comunicación requiere de un lenguaje, lo que debe ser la primera pieza de la construcción de un exocerebro. Quienes se dedican a crear y montar aparatos y sistemas dotados de inteligencia artificial deberían prestar más atención a lo que podría significar una cultura robótica, es decir, el conjunto de símbolos y costumbres que formarían el tejido de las relaciones y comunicaciones entre máquinas, y entre máquinas y humanos.

Ante el posible desarrollo de una cultura robótica no podemos esperar una evolución similar a la que llevó a los humanos primigenios a generar un exocerebro. La gran diferencia es que ahora nos enfrentamos a un proceso guiado y programado intencionalmente por humanos, que aquí son como los dioses creadores de la nueva cultura. La humanidad evolucionó sin la intervención de deidades ni de fuerzas metafísicas, mediante un proceso de selección natural y de mutaciones. El mundo de los robots, en contraste, crece y evoluciona bajo la dirección técnica, la guía intelectual y los intereses de seres humanos que han decidido crear un mundo nuevo para la inteligencia artificial. Es como si un dios hubiese dotado de alma a los robots. No hubo ningún dios en el origen de la cultura humana, pero sí lo hay en el nacimiento de la inteligencia artificial: la sociedad que ha decidido impulsar nuevas formas de trabajo automatizado.

¿Hay un lenguaje que comunica a los robots con su entorno? Las

máquinas inteligentes usan los algoritmos propios del lenguaje de la computación y la lógica matemática. Los ingenieros usan este lenguaje para programarlas. Los robots funcionan con ese lenguaje, pero no lo entienden, de la misma forma en que nuestros cerebros operan con señales eléctricas que (todavía) no comprendemos. La conciencia surge cuando aparece una comunicación entre las señales neuronales y los símbolos culturales. En el caso de las máquinas inteligentes podría ocurrir algo similar: un puente comunicaría dos ámbitos que se basan en configuraciones y lenguajes diferentes. Esta condición híbrida podría ser el origen de la conciencia robótica. Hoy los ingenieros sueñan con la emergencia de una singularidad que aparecería en las máquinas inteligentes del futuro.

Todavía no comprendemos bien cómo emergió en los humanos la singularidad de la conciencia. Desde mi punto de vista, la explicación se encuentra en la evolución y el funcionamiento del sistema híbrido que conjunta las señales cerebrales con los símbolos culturales. La robótica podría ser de gran ayuda para explicar la conciencia humana, pues en su ámbito técnico es posible experimentar de forma práctica y concreta con modelos híbridos y con aparatos operativos. En todo caso, en nuestro vuelo imaginario hacia el futuro, en algún momento nos encontraremos con una cultura robótica que se constituya en el cemento cohesionador de una colectividad de máquinas, un conglomerado del que seguramente formarán parte los humanos, los demiurgos de esta singularidad.

Si observamos el mundo de la computación y de la inteligencia artificial con ojos de antropólogo, podremos ver que hay allí una especie de cultura embrionaria basada en el cálculo, la programación, la estadística, la lógica matemática y la formalidad de las reglas. Se trata de la cultura del procesamiento de información, de las representaciones en forma sistemática y ordenada basadas en las matemáticas de Charles Babbage, el álgebra de George Boole y el mundo de Alan Turing. Es el universo de lo computable y de los algoritmos. Esta cultura informática y cibernética, para que pueda expandirse, requiere de puentes que la ligen al conjunto de la sociedad. Ese fue el sueño de Leibniz, el gran filósofo y matemático del siglo XVII, quien imaginó un lenguaje universal que operase como el álgebra o, mucho mejor, como un conjunto de ideogramas. Le pareció que los caracteres chinos podían servir a su propósito de construir esta especie de codificación del pensamiento a la que llamó *characteristica universalis*. Esta

característica universal debía formarse con signos básicos que representarían nociones cuyas partes ya no pudieran ser separadas para comprenderlas (es decir, nociones no analizables). El lenguaje universal sustituiría los fonogramas del habla por ideogramas precisos y concisos que podrían ser representados por numerales; estos numerales expresarían las verdades científicas fundamentales y podrían ser computados mediante operaciones matemáticas. Ciertamente, en matemáticas los ideogramas son superiores a los fonogramas. El lenguaje universal, creyó Leibniz, debía manejarse mediante lo que llamó el *calculus ratiocinator*, el cálculo de razonamientos.

La idea de Leibniz implica la construcción de una especie de traductor general capaz de transformar las ideas en números. De esta forma las disputas intelectuales se resolverían gracias al cálculo. En la sociedad los símbolos cristalizan en palabras que se acumulan y forman una gran torre de Babel, desordenada y caótica, plagada de diferencias e incoherencias. Al traducir los símbolos a los signos ideográficos de un álgebra global se conseguiría, mediante el cálculo, llegar a verdades universales. Este lenguaje artificial lograría reducir el razonamiento a la computación. Leibniz nunca construyó este lenguaje, que sigue siendo el sueño de los diseñadores de máquinas inteligentes: lograr que sus algoritmos simulen el pensamiento con tal perfección que su comportamiento no se pueda distinguir del de una persona de carne y hueso (es decir, superar la famosa prueba de Turing). Leibniz y sus sucesores han soñado con la invención de un sistema híbrido capaz de juntar el mundo inestable de las razones y los símbolos culturales con el mundo ordenado de los signos matemáticos y el cálculo. De esta manera híbrida funciona nuestra conciencia.

Ante las grandes dificultades teóricas y técnicas para diseñar y construir máquinas inteligentes dotadas de la condición híbrida que he descrito, se han buscado caminos alternativos. La cultura robótica basada en tradiciones computacionales ha adoptado la idea de que los cerebros humanos, al menos en parte, funcionan como computadoras. El pensamiento tendría como soporte un idioma cerebral llamado mentalés que supuestamente funciona de forma similar al lenguaje con que están programadas las computadoras.⁷⁸ Por ello se cree que en principio sería posible traspasar o copiar el contenido mental de un cerebro a una computadora. Esta operación, que implica cargar o subir a una computadora los algoritmos de

una mente, podría producir un sistema que gobernase una máquina de manera similar a como los pensamientos gobiernan nuestro cuerpo. Se trataría de hacer una copia de un cerebro para implantarla en un sustrato no biológico de carácter computacional y mecánico. El proyecto de hacer una copia del cerebro para subirla a una computadora y construir así la réplica de una persona insertada en un robot me recuerda el caso de ese colono inglés que en el siglo XIX quiso copiar la empresa capitalista para implantarla en Australia. Karl Marx se burló de la aventura de este empresario, que le sirvió para explicar el carácter del capital. Este colono, Thomas Peel, se llevó en un barco medios de producción, víveres y obreros con sus familias a Swan River. Pero una vez que se estableció allí se quedó sin ni un sirviente que le hiciera la cama y le subiese agua del río. Los obreros que llevó consigo se dispersaron enseguida, en busca de oportunidades mejores en los vastos territorios poco habitados del oeste australiano. El pobre capitalista había previsto meticulosamente llevar todo lo necesario, pero se olvidó algo fundamental: las relaciones de producción.⁷⁹ Lo mismo le sucedería al ingeniero que lograra copiar todos los detalles de un cerebro para montarlos en una computadora: no se crearía una mente semejante a la original, pues se habría olvidado de trasladar las relaciones sociales, que son parte indispensable de la conciencia. Habría copiado el cerebro, pero se habría olvidado del exocerebro. Para evitar este fracaso hay quienes imaginan, a la manera del novelista Philip K. Dick, que una supercomputadora pudiese albergar a muchas personas viviendo en una sociedad que no tendría más que una existencia digital.

Este tipo de propuestas suelen hacer a un lado la diferencia entre las redes cerebrales operando con señales y las prótesis culturales que se basan en símbolos, pues supuestamente tanto el cerebro como las computadoras inteligentes comparten un lenguaje similar. Desde esta perspectiva no tendríamos un problema para descifrar los procesos de la conciencia, pues la mente humana y las computadoras se regirían por principios comunes. No habría hibridismo. Podemos sospechar que hay un cierto isomorfismo entre las estructuras cerebrales y las culturales. Pero dudo que dichos isomorfismos sean provocados por la coincidencia de estructuras computacionales en los diferentes ámbitos y especialmente entre robots y cerebros humanos. Para explorar este problema desde otro ángulo, será

necesario considerar la relación entre los humanos y las prótesis que han inventado.

7. PRÓTESIS Y SÍMBOLOS

Hay un experimento mental que circula en diferentes variantes desde hace años entre los interesados en la inteligencia artificial. Hay que suponer que a una persona se le sustituye una neurona por una réplica de silicón. Lo más probable es que nadie note un cambio. Pero si se siguen sustituyendo neuronas por prótesis de silicón, ¿habrá un momento en que la persona deje de ser la misma? ¿Su conciencia será la misma con la mitad de las neuronas de su cerebro operando como prótesis de silicón? ¿Tendrá los mismos recuerdos y los mismos gustos? Se parte de la idea de que el comportamiento es el resultado de procesos físicos que pueden ser replicados en prótesis electrónicas de silicón. Así, si todas las neuronas fuesen al final sustituidas por réplicas artificiales, en principio la persona no cambiaría ni se convertiría en un zombi carente de vida interior. Estaría enamorada de su misma pareja, odiaría como siempre al jefe que en su trabajo la maltrata y le seguiría provocando envidia el amigo que la derrota casi siempre en el juego de go. Y si ocurre el proceso inverso, como supone Murray Shanahan, y gradualmente se vuelven a instalar en el cerebro de esa persona neuronas biológicas, en principio no habrá tampoco un cambio en la conciencia. Ello probaría que lo importante es la función de la neurona y no su constitución biológica.⁸⁰

En este experimento mental se parte de la premisa de que la conciencia está contenida en el cerebro. Como en esta prueba se asume que la persona no ha cambiado de ambiente social, se evade el hecho de que una parte de la conciencia en realidad se encuentra en su entorno cultural. A partir de este experimento mental a veces se salta a la propuesta de que la conciencia de una persona podría ser extraída o copiada para subirla a una supercomputadora. Una variante de esta idea es la imitación y emulación de un cerebro biológico en una computadora. Para ello sería necesario generar un mapa extraordinariamente preciso del cerebro humano, diseñar una simulación de la actividad electroquímica que conecta todas las neuronas y construir un soporte corporal sintético que establecería una comunicación con el mundo exterior. Desde luego, hoy no es posible llevar a la práctica

esta idea, pues no existen las técnicas de mapeo, ni la capacidad de programar un símil de la actividad electroquímica, ni la tecnología para construir el soporte para todo ello. La capacidad de cómputo que se necesitaría supera con mucho las posibilidades actuales. No es previsible que en un futuro cercano se pueda programar la actividad de los 86.000 millones de neuronas que emularían un cerebro humano. Ni siquiera es pensable crear un modelo con los más de 70 millones de neuronas del pequeño cerebro de un ratón. Además, la tarea es hoy imposible pues no se conoce la verdadera naturaleza de las señales que se transmiten en el cerebro. El modelo de la comunicación electroquímica entre neuronas no ha logrado descifrar los códigos de las señales. Incluso podría ser que las señales no fuesen electroquímicas sino electromecánicas.

Otra alternativa podría ser la del cibernético. En este caso se sustituirían las diferentes partes del cuerpo por prótesis artificiales de brazos, corazón, estómago y demás partes de la anatomía. Los órganos vivos serían retirados y se sustituirían por prótesis implantadas, hasta dejar solamente el cerebro conectado a un cuerpo totalmente artificial. ¿Se mantendría la identidad de la persona? Se podría ir más allá, para evitar las estorbosas complicaciones técnicas en la construcción de un cuerpo artificial, y crear un cuerpo virtual en una computadora y añadir en ella un entorno social y material simulado. Tendríamos al final un cerebro conservado vivo en un frasco y conectado a redes electrónicas programadas en una computadora. Se supondría en este caso que, por haber conservado una porción biológica de la persona –su cerebro– se mantendría no solo la misma conciencia, sino también la capacidad de sentir emociones. Una variante aún más radical sería la de suponer la posibilidad de llegar a elaborar material biológico sintético modelado como un cerebro al que se le inyectarían los datos de una persona real.

Todas estas alternativas imaginarias están apoyadas en la premisa de que la conciencia humana y el cerebro funcionan esencialmente con algoritmos, como sistemas de información regidos por las reglas de la lógica matemática, la cibernética y la computación. Estas son especulaciones típicas de la investigación en la llamada inteligencia artificial simbólica. Pero aquí el uso del concepto de símbolo es el de las matemáticas que realizan cálculos exactos mediante expresiones que usan variables sin un valor dado. Estas variables y expresiones se manipulan como símbolos en la

lógica matemática, pero no corresponden al concepto antropológico de símbolo que yo uso. Un símbolo es, por ejemplo, una palabra cuyos sonidos (y letras) se asocian a un significado concreto que puede ser un objeto o un concepto, y son una herramienta del pensamiento y la imaginación. Los símbolos matemáticos, en contraste, son abstractos y arbitrarios, sea que representen una función (+, $\sqrt{\quad}$, -, =, x) o una cantidad no definida representada por letras en el álgebra que usan los programadores. Por el contrario, las palabras de una lengua se asocian a significados concretos y si se sustituyen arbitrariamente el resultado será la confusión. Los símbolos en matemáticas son puros y abstractos. En cambio, los símbolos en la lengua, la literatura, la música y las artes tienen cualidades que hacen pensar y generan sentimientos, pero no lo hacen de una manera exacta e idéntica en cada persona y en todo momento. Estos símbolos no pueden ser manipulados por el lenguaje computacional que usa unidades de información. Ya hace mucho que la inteligencia artificial simbólica fue criticada por creer erróneamente que el cerebro es como el hardware y la mente el software de las computadoras. La mente no opera mediante reglas algorítmicas basadas en la simbología matemática. Por estas razones, prefiero hablar de signos matemáticos y no de símbolos.

Hoy en día la robótica no puede abandonar el uso de la signología lógico-matemática para programar aparatos inteligentes. Se enfrenta a un problema similar al que nos topamos cuando intentamos comprender el funcionamiento de la conciencia humana, en la que se juntan en el mismo sistema las señales electroquímicas y los símbolos culturales. Un robot inteligente tiene que funcionar con un sistema que opera con signos de carácter matemático y señales electromecánicas en un ambiente basado en símbolos culturales. No es muy útil suponer que la mente humana opera con algoritmos y procesos matemáticos a semejanza de los robots. Los aparatos inteligentes de alto nivel tendrán que resolver el problema de la interacción entre sistemas matemáticos y redes simbólicas de carácter cultural. La inteligencia artificial, si se quiere que tenga un nivel similar a la humana, deberá basarse en un sistema híbrido. Como he dicho, deberá formarse un exocerebro, que puede tener un carácter mecánico o humano. Si el exocerebro fuese humano, podría pensarse que alguna de las prótesis artificiales inteligentes que usamos se desarrollaría a tal grado que

convertiría a sus creadores en una prótesis, invirtiéndose los papeles originales.

Si el exocerebro es un aparato inteligente y también mecánico, podría ser la extensión de lo que anteriormente era una prótesis y que se ha vuelto autónoma; es decir, estaríamos ante la prótesis de una prótesis. Esta situación podría implicar una larga cadena de aparatos, cada uno formando un exocerebro del eslabón anterior. De forma más compleja y sofisticada, podría ser un conjunto de máquinas no eslabonadas en serie sino funcionando en paralelo, pero interconectadas. Un sistema de prótesis de esa naturaleza funcionaría de tal manera que cuando una de las máquinas inteligentes tropezase con un problema y quedase atascada, pudiese acudir a otros aparatos para superar el atolladero (u obtener la ayuda de ingenieros humanos). ¿Cómo podemos saber si en este acto de recurrir hay una voluntad libre consciente y no una mera dependencia técnica?

Para contestar esta pregunta se han buscado explicaciones diversas. La más socorrida es la que explica que si el aparato inteligente *parece* ser consciente, es que realmente lo es. Se cree que eso es lo que les sucede a los humanos. Otra explicación, presentada por Michael Graziano, es la llamada teoría del esquema de atención (*attention schema theory, AST*, en inglés).⁸¹ Esta teoría considera que nuestro cerebro es un procesador biológico de información que llega a *pretender* que posee una atención subjetiva, no física, que se percata del entorno.

Este mismo proceso podrá operar en un futuro no muy lejano, cree Graziano, cuando aparezca la conciencia artificial. En ese futuro será fácil la transferencia de las mentes a un nuevo hardware, a la nueva maquinaria robótica. Estos aparatos inteligentes podrán estar dotados, como los humanos, de un esquema de atención que construye un modelo del cuerpo y de los procesos internos. Esta conclusión es similar a la de quienes creen que la conciencia es una ilusión. Pero Graziano no cree que el esquema de atención experimente una ilusión: no experimenta *nada*, no tiene ninguna experiencia subjetiva. Una máquina dotada de este esquema tiene una información falsa o simplificada que le dice que tiene experiencias. Igualmente, se supone que los humanos creemos tener experiencias subjetivas porque hemos aprendido socialmente la idea ingenua de que las

tenemos. Una máquina inteligente debe contener información interna que le indique que tiene conciencia. Así, los ingenieros deberían poder construir una máquina que crea que es consciente, que pretenda serlo y que hable de la conciencia de la misma manera en que lo hace la «máquina humana». Pero, añade, debe estar basada en la inteligencia social, pues sin ella las computadoras serían sociópatas, lo que sería muy peligroso.

Estas ideas sobre la conciencia artificial se basan en una especie de degradación de la experiencia subjetiva de los humanos. Al rebajar el nivel de la conciencia humana se produce el efecto de que las máquinas nos pueden alcanzar pronto y exitosamente. Se trata de un espejismo que puede ser útil para estimular a los ingenieros constructores de robots y para descubrir mecanismos originales nuevos con que dotar a los aparatos. Pero la verdadera conciencia artificial continuará eludiéndonos.

8. EXPERIENCIAS ROBÓTICAS

El ingeniero japonés Jun Tani, en su trabajo de construcción de robots, se guía por la famosa idea del gran físico Richard Feynman según la cual no podemos entender aquello que no podemos crear. Siguiendo esta idea, Tani sostiene que es posible entender lo que se puede crear. Para entender la conciencia se ha propuesto crearla artificialmente. Su propuesta, muy inteligente y creativa, es aplicada a experiencias con robots relativamente elementales pero sensibles a su entorno, capaces de circular en espacios cerrados, evadir obstáculos y perseguir el objetivo de llegar a un lugar preciso. Desde luego, parte de la elaboración de modelos matemáticos para incorporarlos a sus robots.⁸² Se trata de redes neuronales recurrentes (RNR, en inglés: *RNN, recurrent neuronal networks*) creadas como un modelo conexionista capaz de manejar estructuras tanto espaciales como temporales gracias a sus propiedades dinámicas. Tani parte de la idea de que el modelo neuronal debe abarcar la interacción entre las intenciones, que van de arriba abajo, el reconocimiento perceptual del entorno, que procede de abajo hacia arriba. Desde luego, en este esquema, arriba se encuentra la mente subjetiva y abajo el mundo objetivo. Este sistema funciona con signos matemáticos, mientras que el cerebro humano opera con señales electroquímicas. En las redes que funcionan como el cerebro del robot se forma una causalidad circular dominada por la dinámica no lineal que conecta las intenciones con la percepción de la realidad. Se trata de la aplicación de la teoría del caos a la construcción de redes neuronales robóticas. Para ello, el sistema dinámico de redes neuronales robóticas debe generar lo que se denomina un «atractor extraño». Para que esto ocurra, el sistema debe ser sensible a las condiciones iniciales, de tal manera que los acontecimientos son atraídos hacia trayectorias cercanas que convergen de manera espontánea. Hay una extraña confluencia de un efecto determinista, causado por la dependencia respecto a las condiciones iniciales, con la formación de convergencias que atraen en forma caótica los acontecimientos hacia formas impredecibles. Estos atractores combinan dependencia e independencia. Su independencia solo puede entenderse (y predecirse) mediante el cálculo de probabilidades,

pero al mismo tiempo se trata de un fenómeno dependiente causalmente de las condiciones iniciales.

Los robots dotados de estos sistemas dinámicos no lineales nunca actúan reactivamente ante los estímulos externos. Actúan guiados por mapas de conducta generados por la relación entre las intenciones y las percepciones, es decir, por un sistema que aúna en un flujo vectorial tanto las intenciones como las percepciones. Este flujo forma un atractor estructuralmente estable que es el responsable de las trayectorias y acciones del robot. Una imagen sencilla puede dar una idea del funcionamiento de la conciencia del robot. Su mente sería como la espesa columna de humo que sale de la chimenea de una fábrica: siempre es la misma pero siempre es diferente, pues sus formas cambian espontáneamente según los vientos, la humedad, la temperatura y muchos otros factores. Pero en el robot la columna de humo tiene funciones cognitivas y ejecutivas, genera patrones espaciotemporales continuos y mantiene una sensibilidad ante el entorno.

Así, una dinámica caótica gobierna al robot y puede producir lo que Tani llama «combinatorias de símbolos». Lo logra dividiendo el flujo de la acción en trozos, cada uno de los cuales «representa» cosas. El «pensamiento» del robot es segmentado en operaciones lógicas que funcionan como combinaciones de símbolos. Lo más importante es que al confrontarse las intenciones con la experiencia que llega de los sensores se emite una señal de error y se modifica el comportamiento. Tani cree que la señal de error provoca la experiencia en el robot de una conciencia que se percata de su subjetividad. El robot percibe que la intención subjetiva difiere de la realidad objetiva circundante y la máquina se siente «fuera de lugar», una condición que no deriva de la autoproyección. Cuando se percibe un error, el flujo de sensaciones se segmenta por medio de una modificación paramétrica vectorial para minimizar el conflicto. Allí surge la conciencia, cree Tani.

Durante el entrenamiento del robot Yamabico, dotado de un modelo RNR, aparece en sus redes neuronales, en un momento dado, una región altamente caótica, un extraño atractor. En la transición de una condición estable a otra inestable aparecen primero señales de error en la predicción, pues hay conflictos entre la expectativa en la mente subjetiva del robot y lo que sucede al actuar en la realidad que lo rodea. En este momento de «incoherencia» aparece la «autoconciencia» del robot, pues el sistema se

dirige a resolver el conflicto. En contraste, durante la fase estable se reduce la autoconciencia, pues no hay conflictos que exijan la atención del sistema.

El sistema no opera, propiamente hablando, con símbolos sino con estructuras autoorganizadas en un sistema neurodinámico (RNR) que de manera natural «ateriza» en el entorno físico del robot, permitiendo una interacción entre el sistema y el espacio real. Así aparece espontánea y momentáneamente un yo robótico. Al experimentar con el robot bípedo QRIO de Sony, dotado de una red neuronal recurrente de escala temporal múltiple (MTRNN), el aparato se vuelve «consciente» al reconocer los estados mentales de otros robots cuando alterna con ellos mediante actos motores. En los experimentos con otro robot humanoide de Sony se observó que durante su entrenamiento las secuencias comenzaron a desviarse de las pautas aprendidas debido a la sensibilidad a las condiciones iniciales del sistema. Esta aparente espontaneidad podría deberse al ruido del mundo físico exterior. Sin embargo, cuando se eliminó el ruido externo también se originaron acciones espontáneas procedentes del interior del sistema dinámico. La subjetividad, cree Tani, no es un estado sino una función dinámica en la predicción de los resultados percibidos que resultan de la interacción con el mundo objetivo.

Los modelos diseñados por Tani utilizan las ideas de los neurocientíficos que han postulado que la mente humana funciona como un sistema dinámico que genera atractores caóticos. El neurólogo Walter Freeman, por ejemplo, estaba convencido de que la intencionalidad es espontáneamente generada mediante el caos en la corteza prefrontal. Tengo mis dudas de que el cerebro humano funcione de acuerdo con sistemas dinámicos que generan extraños atractores caóticos. Pero sin duda la programación de estos sistemas es la que puede acercar a un robot a un comportamiento similar al humano. Tani reconoce que la espontaneidad y la voluntad con que operan estos robots no son generadas libremente, pues surgen de la causalidad determinista de los estados internos. Cree que desde el punto de vista matemático el libre albedrío total, sin ninguna causalidad previa, podría no existir. Acaso lo que no existe es el azar profundo, pero no debemos confundir la libertad con el comportamiento aleatorio. Los humanos llegan a ser libres no porque actúan al azar sino porque son capaces de una singularidad: interrumpir la cadena causal. Desde el punto

de vista de la teoría del caos, aquí es donde entra la sensibilidad a las condiciones iniciales.

En este punto llegamos al enfrentamiento entre la dinámica determinista y los procesos probabilísticos en la generación de la espontaneidad en robots. En la construcción de robots Tani prefiere el modelo determinista en los sistemas dinámicos al uso de secuencias probabilísticas, pues los modelos probabilísticos se paralizan cuando hay una contradicción entre las intenciones y la información externa. En cambio, el modelo determinista puede evitar la parálisis al ser capaz de desarrollarse con autonomía. Al usar un mecanismo de regresión de errores, el sistema desarrolla caos durante el entrenamiento y con ello genera nuevas acciones al tratar de evitar caer en los patrones habituales de comportamiento. Al ver que las respuestas externas contradicen las expectativas, el robot genera una señal de error en su sistema y en consecuencia su estado neuronal es modificado por medio de un proceso de regresión de errores.

De forma muy primitiva, estos robots operan con una predicción del futuro y una posdicción del pasado, pues las acciones propias son recogidas en forma posdictiva cuando las intenciones iniciales son reformuladas para adaptarlas a la realidad. Así pues, la conciencia según Tani es la sensación de que la estructura neuronal se modifica para adaptarse a un entorno cambiante imprevisto o imprevisible. Me parece un gran acierto de Tani la propuesta de que el sistema operativo de un robot debe incluir su medio ambiente. Comprende bien que el cerebro humano no funciona con símbolos, pues estos se encuentran en el entorno. Como buen ingeniero, se da cuenta de que la conciencia de un robot, que debe plasmarse en un sistema, tiene que formularse como un modelo matemático que pueda incorporarse a la máquina. Lo más cercano a la conciencia humana que ha encontrado es el sistema dinámico no lineal de atractores caóticos. Pero su idea de que así funcionan los cerebros humanos no parece sólida. Tani es capaz de crear robots en los que parece surgir la chispa de la conciencia de una forma muy elemental y primaria. Pero, así como reconoce que en sus robots no hay un verdadero libre albedrío, podemos también observar que no están dotados de una auténtica conciencia como la humana. Pero es, por ahora, un buen acercamiento a la construcción de una conciencia artificial propia de los robots y regida por el determinismo dinámico de los sistemas caóticos.

9. LA EMANCIPACIÓN DE LOS EXOCEREBROS

La naturaleza ha creado algo tan extraordinariamente antinatural como la cultura, ese conjunto de artificios y artefactos que definen a los humanos. La conciencia emergió como un fenómeno montado tanto sobre la naturaleza biológica de las personas como sobre las estructuras simbólicas artificiales y los instrumentos a ellas ligados. Nos podemos preguntar si algunos de los artificios que forman parte de la conciencia podrían evolucionar hasta separarse de los seres humanos, independizarse y emanciparse de su parte biológica. El lenguaje es la pieza exocerebral más importante de la conciencia humana. Podemos comprender que una cristalización del lenguaje en, por ejemplo, un libro forma parte de nuestra conciencia. El libro adopta la forma de una realidad inorgánica que puede aparecer como un conjunto de hojas de papel cosidas, como una secuencia de símbolos digitales albergado por una computadora o por un teléfono inteligente. El libro forma parte de la conciencia solamente cuando se conecta con el cerebro humano. En el momento en que se conecta con el lector, el libro pareciera despertar y cobrar vida. Entonces se funden el mundo cerebral y el mundo escrito en una sola estructura. ¿Qué puede hacer consciente a una prótesis como la computadora que alberga el libro que leemos? Su conexión con el cerebro. Lo mismo ocurre a la inversa: el cerebro es autoconsciente solo cuando se conecta a una prótesis que lo comunica con los humanos de su entorno. La conciencia humana es el conjunto endo y exocerebral. El cerebro sin el habla no puede ser autoconsciente. Cuando el cerebro se conecta con un objeto que forma parte de su entorno simbólico, este objeto se convierte en parte del sujeto.

Lo que no se ha logrado realizar es la transformación de un objeto en sujeto sin la intervención de un cerebro humano. Las máquinas inteligentes, como las conocemos hoy, son objetos y solo se convierten en sujetos si se conectan con cerebros biológicos. ¿Podemos suponer que en algún momento puede iniciarse una emancipación de las máquinas inteligentes, de tal manera que dejen de ser objetos para convertirse en sujetos? Este salto

singular solo puede ocurrir si, como ya he señalado, el artefacto inteligente desarrolla un exocerebro, sea mecánico o biológico. Una conciencia artificial similar a la humana no puede funcionar solamente basada en un sistema dinámico que genera atractores caóticos capaces de simular el libre albedrío propio de la voluntad humana. El sistema interno de una máquina inteligente debe formar parte de una red que incluya el equivalente cibernético de un exocerebro humano. O bien que, en un singular intercambio de papeles, los cerebros humanos acaben siendo las prótesis de unos robots superinteligentes que no solo se han emancipado sino que han sometido a los humanos a una simbiosis.

De cualquier manera, los ingenieros se enfrentan a un problema parecido al que confronta a los neurocientíficos: comprender la manera en que las señales electroquímicas neuronales se conectan con la red de símbolos exocerebrales. En el caso de los robots es necesario encontrar la forma de integrar un sistema de signos lógico-matemáticos con los símbolos propios de un exocerebro. Por un lado, tenemos la dimensión neuronal o mecánica que opera con señales y signos; en el otro lado, pero estrechamente conectada, tenemos la dimensión cultural simbólica.

Las neurociencias no han logrado todavía descifrar los códigos con que funciona el sistema nervioso central. El estudio de los procesos químicos y eléctricos no ha descubierto cómo opera el sistema neuronal. Se supone que las imágenes y las palabras son procesadas por las proteínas que en las neuronas abren y cierran canales iónicos en las sinapsis, lo que genera cambios en el voltaje que polarizan y despolarizan las membranas. Pero esta actividad de los llamados neurotransmisores no se ha descifrado y por lo tanto no se entiende cómo los procesos químicos y eléctricos procesan el pensamiento o los sentimientos. El hecho es que las neurociencias no nos han proporcionado una teoría general sobre cómo funcionan los circuitos neuronales, cómo generan estados mentales o comportamientos, o cómo se producen desequilibrios mentales. El estudio de las neuronas hasta ahora ha fracasado y se ha visto la necesidad de ampliar el horizonte de las investigaciones (por ejemplo, estudiando conjuntos de neuronas más que la operación de cada una de ellas).

Es muy probable que el modelo de transmisiones neuronales sea incompleto o incluso equivocado. Es lo que piensa el físico Thomas Heimburg, cuyas investigaciones sobre las membranas biológicas con un

enfoque termodinámico lo han llevado a afirmar que la propagación de las pulsaciones nerviosas no puede ser explicada por los procesos electroquímicos basados en el viejo modelo de Hodgkin-Huxley, establecido hace más de sesenta años. Heimburg sostiene que las señales nerviosas se propagan mediante pulsaciones mecánicas bajo la forma de oleadas, como los llamados solitones, esas ondas solitarias que se propagan en medios no lineales por largos trechos sin cambiar. La transmisión de señales en las membranas nerviosas sería más bien un fenómeno electromecánico.⁸³ No sé si por esta vía se llegarán a encontrar las claves cerebrales del pensamiento y a descifrarse las señales que transitan por el sistema nervioso central. Pero esta nueva interpretación es posiblemente un síntoma de que el paradigma neuronal está estancado en un callejón sin salida.

En contraste con lo poco que se sabe sobre la transmisión de las señales en el sistema nervioso, los ingenieros entienden muy bien el funcionamiento de los «cerebros» de los robots, pues ellos mismos los han diseñado y construido. Pero de momento no se sabe cómo las máquinas inteligentes podrían desarrollar un sistema simbólico que les permita realizar las funciones y operaciones que su cerebro electrónico no puede realizar. De hecho, los constructores de robots carecen de una teoría general del sistema cognitivo universalmente aceptada, de manera que tienen que explorar alternativas basados en diferentes interpretaciones. Yo estoy reflexionando a partir de la teoría que he desarrollado sobre el exocerebro, pero son pocos los diseñadores de robots que se han propuesto programar el sistema operativo de sus máquinas incluyendo un entorno simbólico complejo que sirva de apoyo y de prótesis. Aquí estoy pensando en unos robots dotados de un sistema enteramente nuevo que no sea ni una copia del cerebro humano ni una estructura similar a las redes neuronales. La otra opción sería la de cargar el contenido de un cerebro a una computadora, una solución meramente imaginaria, pues no solamente no existe hoy la tecnología para realizarla, sino que, como he dicho, los científicos no saben cómo la actividad cerebral genera la mente. No es posible copiar o cargar el contenido de un cerebro si ignoramos el sistema de señales con que opera.

Es preferible, por ahora, partir de los robots y sistemas realmente existentes para especular sobre la manera en que podrían desarrollar una conciencia artificial. Si el cerebro de un robot, como los que ha descrito Jun

Tani, contiene un sistema dinámico no lineal, como un extraño atractor, es necesario que sea capaz de entender y procesar los símbolos de su entorno, además de rodar sin chocar con las paredes y dirigirse a objetivos determinados. Estos símbolos son, por ahora, los que impulsan e ilustran a los humanos que rodean al robot. Los ingenieros tienen que encontrar una forma en la que puedan confluír en su sistema los signos numéricos y los códigos simbólicos. Ello requiere que los robots posean una inteligencia flexible y amplia no muy especializada capaz de enfrentar múltiples tareas.

Este problema no se puede enfrentar bien si se parte de la idea tan generalizada hoy de que las mismas reglas matemáticas se aplican tanto a los organismos humanos como a las máquinas electrónicas. Se parte con frecuencia de la idea equivocada de que todo organismo funciona mediante algoritmos y que las actividades animales (incluyendo a los humanos) están reguladas por mecanismos de procesamiento de datos. Como ya he señalado, las computadoras electrónicas y los cerebros biológicos no tienen en común algoritmos reguladores de flujos e información codificados mediante signos matemáticos o lógicos.

Estamos ante un problema similar al que enfrenta la física: el mundo macroscópico no funciona con las mismas reglas que el mundo microscópico. Este último se rige por la teoría cuántica, de naturaleza probabilística; en cambio el otro mundo funciona con las reglas de la física clásica newtoniana. Pero en realidad los dos mundos coinciden uno dentro de otro de una manera que los científicos no entienden todavía. Algo similar ocurre con el fenómeno de la conciencia; allí confluyen dos esferas, el mundo biológico del cerebro y el mundo simbólico del exocerebro. El primero opera mediante procesos mecánicos, eléctricos y químicos, y el segundo funciona con un sistema simbólico de índole cultural y social. La reducción de un mundo a las normas del otro no permite entender el fenómeno de la conciencia. Pero aquí estamos en una situación peor que la de los físicos. La mecánica cuántica es bien comprendida y ello ha impulsado decididamente a la industria electrónica moderna. Lo que no entienden los físicos es el momento en que los dos mundos se tocan, por llamarlo así, y ocurre lo que Werner Heisenberg llamó un «colapso». Ello ocurre en el momento en que un humano ejecuta la medición de una onda, acto que precipita que la partícula «escoja» una posición o un estado. Nadie entiende por qué la intervención humana consciente provoca que se esfume

la superposición cuántica. La teoría explica lo que *podría* suceder, pero no la mecánica de lo que sucederá cuando se transita de lo probable a lo real. El colapso cuántico ha generado otras interpretaciones, como la idea de la localización espontánea continua (CSL), según la cual el colapso ocurre de manera azarosa. También se ha visto la pérdida de la coherencia como explicación del colapso. Otra interpretación plantea que no hay colapso ni decoherencia, sino que todas las posibilidades son reales y coexisten en un multiverso. Ya me he referido más arriba a la paradoja del gato de Schrödinger para ejemplificar la forma en que la física introduce el tema de la conciencia humana en la teoría cuántica.

Cuando el mundo cerebral entra en contacto con el mundo cultural hay una especie de «colapso» y lo que funciona de acuerdo con las leyes biológicas deterministas tiene consecuencias sociales, políticas y económicas que se rigen por reglas muy diferentes. Una diferencia muy importante entre ambos conjuntos de reglas es que en el universo biológico no hay libre albedrío, pero en el momento en que se conecta con el mundo cultural hay un «colapso», pues surge súbitamente la libre voluntad como una capacidad propia de los humanos. La manera fácil de escapar de esta paradoja es decretar que los humanos no tienen un yo único capaz de decidir y elegir. El libre albedrío sería un mito. Esta evasión permite cerrar los ojos ante el hecho de que la conciencia humana es un fenómeno que evade el determinismo científico tradicional.

¿Podrán los ingenieros conectar el cerebro de los robots a un sistema externo de carácter simbólico que no funcione con algoritmos matemáticos? Se podría experimentar probando dos vías. La más remota consiste en la organización de una sociedad de robots interconectados que sean capaces de generar de manera autónoma una red de vínculos culturales, instituciones y convenciones simbólicas. Esta alternativa requiere el diseño de robots con una inteligencia general muy desarrollada, algo lejano que todavía no se avizora. La segunda alternativa consiste en conectar a robots inteligentes, aunque todavía especializados, con las redes sociales y culturales de los humanos. Esta alternativa es algo que ocurre de manera extensa y cotidiana en los ambientes dotados de un desarrollo tecnológico avanzado, donde los humanos están rodeados de robots que les ayudan en muy diversas tareas. Hay que pensar en los asistentes personales como Siri de Apple, Cortana de Microsoft, Bixby de Samsung o Assistant de Google. Pero no vemos en

estos sistemas ni la más mínima chispa de conciencia; en realidad, son más bien prótesis que prolongan la conciencia humana, como lo hacen los teléfonos inteligentes y los diversos programas de traducción o de guía para los conductores de autos. Los carros automáticos sin conductor son también prótesis inteligentes muy útiles, pero cuando uno de ellos atropelló y mató a una peatona en Arizona en marzo de 2018 no parece que haya tenido remordimientos de conciencia.

Si saltamos al mundo del arte, donde hace mucho los robots han intervenido, podemos observar creaciones de gran belleza y originalidad. Jean Tinguely, gran precursor del arte robótico, quería, como dijo, «fabricar máquinas verdaderamente alegres, y por alegres quiero decir felices». Sus Méta-Matics son robots que dibujan obras abstractas mediante un motor eléctrico que mueve tanto el cartón como el brazo que sostiene la pluma. En 1959 su Méta-Matic n.º 17 asombró al público que asistió a la Biennale de París al producir unas cuarenta mil obras en pocos días, todas diferentes y muy atractivas. No creo que esta máquina fuese feliz y su aparente libertad provenía de los movimientos mecánicos azarosos; pero sin duda produjo gran alegría en quienes contemplaron el espectáculo de este primitivo robot artista.

Antes, en 1956, Nicolas Schöffer fabricó la máquina CYSP 1, que fue la primera escultura cibernética en la historia del arte. La escultura contiene un cerebro electrónico conectado a receptores de luz y sonido. La máquina reacciona a estímulos externos y en consecuencia mueve las 16 placas de color que dan vueltas a velocidades diferentes. Sin duda produce un placer estético en los espectadores.

Un ejemplo especialmente interesante es el de Patrick Tresset, un artista que desde 2004 comenzó a crear robots para dibujar retratos de personas con un estilo muy peculiar y atractivo. Inicialmente Tresset dibujaba él mismo los retratos; pero una enfermedad mental –el síndrome bipolar– le afectó profundamente su vida. Al tomar fármacos para controlar su oscilación maniaco-depresiva perdió la pasión por el arte. Como no quería vivir prisionero de la bipolaridad al dejar de medicarse, aplicó sus habilidades en la computación a un robot llamado Paul, dotado de un ojo y un brazo capaz de dibujar en un estilo similar al suyo. En realidad, Tresset creó una prótesis dotada de cierta autonomía que dibuja retratos con una peculiar torpeza o torpeza robótica, como él la llama.

Durante los últimos años se han creado decenas de artistas robóticos que producen una obra impresionante. Una exposición en el Grand Palais de París, *Artistes & Robots* (2018), presentó una muy buena selección de máquinas y de obras que revelan la expansión de un campo nuevo y muy provocador de experimentación artística. Las obras de estas máquinas, sin embargo, no han superado una paradoja: las creaciones de estos robots nos emocionan y nos despiertan sentimientos muy diversos, pero ellos son tan fríos como un témpano de hielo.

10. MÁQUINAS SENTIMENTALES

Yo creo que la singularidad que ha de generar una conciencia artificial está muy lejos. Pero para el inteligente e ingenioso profeta poshumanista, Ray Kurzweil, los primeros indicios de la singularidad se podrán ya ver hacia el año 2045.⁸⁴ Para pronosticar este desenlace tan cercano Kurzweil no se basa en los logros técnicos actuales, sino en una supuesta ley de aceleración de resultados de la investigación, según la cual los descubrimientos crecerían a un fabuloso ritmo exponencial. Ante esta velocidad de los avances técnicos y científicos, asume que la singularidad deberá surgir muy pronto, aunque no se sabe de dónde. Supone que la inteligencia artificial crece a tal ritmo que las máquinas nos permitirán pronto liberarnos del cuerpo biológico. Por supuesto, al igual que la mayor parte de los ingenieros que crean inteligencias artificiales, cree que el cerebro humano y las máquinas comparten algoritmos y mecanismos computacionales que serán la base del singular salto. Ya lo había expuesto uno de los fundadores de la inteligencia artificial, Marvin Minsky: el cerebro es una máquina de carne, pero una máquina a fin de cuentas. Por ello se cree que se podrá descargar el contenido mental del cerebro en una computadora.

Para explorar esta dimensión que parece fantástica quiero viajar a un futuro lejantisimo con la ayuda de un hermoso texto publicado en 1988 por Jean-François Lyotard.⁸⁵ El filósofo francés afirmó que el único problema verdaderamente serio al que se enfrenta la humanidad no son las guerras, los conflictos, las tensiones políticas, los cambios de opinión, los debates filosóficos o las pasiones. El gran problema es que el Sol está envejeciendo y que dentro de 4.500 millones de años explotará, con lo que se acabará todo rastro de nuestro sistema planetario y de la vida en la Tierra.⁸⁶ La única escapatoria, afirma Lyotard, será encontrar la manera de simular las condiciones que permitan que el pensamiento sea posible materialmente cuando hayan desaparecido las condiciones que sostienen la vida en la Tierra. La solución consiste en manufacturar un hardware capaz de «nutrir» un software tan complejo como el cerebro humano actual. Esto es

precisamente lo que se proponen hacer los constructores de inteligencias artificiales y de robots. Pero el pensamiento humano no funciona de manera binaria ni con bits de información: opera de manera analógica y no en forma lógica. El reto consiste en crear algo mejor que un pobre fantasma binarizado de lo que fue antes el cerebro.

Lyotard agrega un problema más, mucho más preocupante: el hecho de que el pensamiento y el sufrimiento se traslapan. Los matices y los timbres cuando se pinta o se compone música o las palabras y las frases cuando se escribe nos son dados pero al mismo tiempo se nos escapan entre los dedos. Todo ello «dice» algo diferente a lo que quisimos «significar», forma parte de un mundo opaco, lleno de vaciedades, que produce sufrimiento. Recuerda que el pensamiento tiene que estar inscrito en la cultura. Lyotard se pregunta si las máquinas pensantes sufrirán o solamente tendrán memorias. Yo agregaría que, además del sufrimiento, el pensamiento va acompañado del placer y de otras emociones.

Hay una dimensión fundamental que preocupa a Lyotard y que rara vez toman en cuenta quienes se dedican a la inteligencia artificial. El cuerpo humano manifiesta diferencias sexuales y de género. El filósofo cree que hay una incompletitud no solo en los cuerpos sino también en las mentes. La incompletitud de la femineidad y la masculinidad no solo es corporal, radica también en el pensamiento, y se intenta superar mediante la atracción, gracias a la poderosa fuerza del deseo. Lyotard se pregunta si la inteligencia que se prepara para sobrevivir a la explosión solar conllevaría esta fuerza dentro de sí en su viaje interestelar. Las máquinas pensantes tendrán que nutrirse no solo de radiación sino del irremediable diferendo del género. El pensamiento postsolar debe estar preparado ante la inevitabilidad y la complejidad de la diferencia y de la separación sexual si quiere sobrevivir a la amenaza de la entropía.

Se dirá que, siendo la humanidad tan joven, pues tiene apenas unos cien mil años de edad, los más de cuatro mil millones de años que nos restan serán más que suficientes para inventar una alternativa técnica que nos libere del cuerpo terrenal que depende de la luz solar y que nos lleve a un viaje cósmico hacia nuevas condiciones de existencia. Sin embargo, la necesidad de desprendernos del cuerpo podría llegar antes que la explosión solar, si por ejemplo ocurriese una maligna mutación de un virus que

obligase a los humanos a mudar su pensamiento del soporte biológico que lo ha albergado desde los orígenes.

Todo esto es una inquietante ficción científica, lejana pero estimulante. Sin embargo, los problemas planteados por Lyotard son reales: los constructores de inteligencias artificiales se enfrentan al reto de introducir en sus programas y en sus máquinas el enigma de los sentimientos y el fenómeno de la sexualidad. Pero estos son unos problemas muy difíciles que rara vez son abordados.

La idea de máquinas sensibles ha sido muy explorada por la ciencia ficción y generalmente se expresa bajo la forma de alguna técnica capaz de cargar la mente que se aloja en el cerebro a una computadora. Es el sueño de los transhumanistas que aspiran a que los humanos se liberen de su soporte blando y húmedo, hagan a un lado la carne, para insertarse en un mecanismo duro y seco que no requiera de ningún proceso biológico para obtener energía. Desde luego, se suele pensar en máquinas emocionales que, a pesar de su sequedad y dureza, serían capaces de gozar y sufrir gracias a esa sensibilidad que nos hace humanos, y que al mismo tiempo nos enlaza con el resto de los animales.

Algunos transhumanistas llevan colgado en el cuello un medallón que recuerda esos colgijes o amuletos en los que se interesó el médico Qusta ibn Luqa. Este medallón identifica al portador como miembro de una institución criogénica dedicada a mantener cadáveres a muy bajas temperaturas para revivirlos cuando se descubra un método para curar la enfermedad que provocó sus muertes, o bien en espera de que algún día aparezca una tecnología capaz de copiar el cerebro muerto y cargar la información a una computadora. El medallón contiene la inscripción no de un conjuro sino de las instrucciones precisas para tratar el cuerpo en caso de muerte y enviarlo a la institución que lo guardará congelado durante siglos. En algunos casos solamente se conserva congelada la cabeza, en espera de que algún día se pueda descargar el contenido del cerebro a una máquina que le proporcione una nueva vida cibernética a la persona. No parece que la singularidad que permita esta operación ocurra pronto, pero sin duda es una alternativa ante el reto que describe Lyotard: para lograr esta hazaña la humanidad tiene mucho tiempo antes que dentro de cuatro millones y medio de milenios desaparezca el sistema solar. Claro que es muy probable que la singular descarga de una conciencia en una máquina no se pueda

realizar más que con seres vivos, por lo que los cuerpos que esperan congelados su resurrección no cumplirán las esperanzas de los transhumanistas que optaron por la criogenia.

Me parece que las máquinas inteligentes no llegarán a tener una conciencia similar a la humana sin tener alguna clase de sensaciones, sentimientos y emociones. Una inteligencia artificial, sea que su conciencia provenga de una descarga del cerebro de un ser humano o que haya sido construida a partir de un diseño nuevo, no podrá funcionar a un nivel como el humano si carece de sensibilidad. Para enfrentarse al reto planteado por Lyotard, estas conciencias inteligentes y sensibles deberán estar montadas en unos soportes no orgánicos que poblarían grandes máquinas espaciales, acaso circulando en órbitas similares a las de un cometa. Pero surgiría un problema que no pensó Lyotard: por su propia estructura metaloide estos seres podrían ser inmortales o vivir muchos siglos o milenios, lo que significaría un salto que modificaría esencialmente la naturaleza humana.

Pero aquí dejaré que el lector imagine un futuro semejante, para retornar al tema de la sensibilidad de las máquinas robóticas. ¿Cómo pueden los ingenieros introducir las emociones y la sexualidad en los robots inteligentes, para que logren tener una conciencia artificial semejante a la humana? Ya cité la idea del físico Max Tegmark, quien imagina una sustancia llamada *sentronium*, a la que ya me he referido. Pero no nos da ninguna idea concreta sobre cómo elaborar o como puede aparecer esa sustancia capaz de tener experiencias subjetivas. El *sentronium* sería como una onda o un algoritmo, que son independientes de su sustrato físico específico. Por lo pronto esta idea es tan inservible como pensar que un soplo divino dota a un robot de sentimientos. La mayor parte de las especulaciones sobre el futuro de la inteligencia artificial parten del supuesto de que este soplo divino (o singularidad, como prefieren llamarla) ocurrirá próximamente.

Ya he insinuado más arriba que podría ocurrir una inversión que acabase convirtiendo a los humanos en prótesis de máquinas superinteligentes. Los humanos serán el *sentronium* de un robot. Esta nueva condición humana podría terminar en la pesadilla de convertirnos en meros pedazos de carne encargados de transmitir sensaciones y sentimientos a poderosas máquinas muchísimo más inteligentes que nosotros. Podemos acaso sentir e imaginar esta nueva condición si recordamos la experiencia del uso de artefactos

cognitivos que apoyan, aumentan e incluso sustituyen las percepciones de los humanos. Es el caso de muchos aparatos, como los generadores de realidad virtual y de ambientes simulados. En algunos casos se trata de juegos y en otros de sistemas de aprendizaje o de observación. Los simuladores de vuelo, como juegos y como sistemas de entrenamiento de pilotos, son un buen ejemplo de articulación de un ser de carne y hueso con una realidad virtual inteligente. Quien haya experimentado con estos simuladores puede imaginar que, de repente, la máquina los convierte en sus prótesis sensibles. Cuando nos conectamos a Second Life, un juego que ha creado una inteligente realidad virtual en la que los humanos se introducen por medio de avatares, podemos aparentemente inyectar nuestras emociones y deseos a un ambiente artificial. Nos hemos convertido en un trozo de carne consciente y sensible enchufado a un mundo artificial por medio de una computadora. Algunos sistemas cibernéticos inteligentes aumentan nuestras percepciones, como los aparatos que registran dimensiones que no percibe el conductor de un auto o el piloto de un avión. Otros sistemas de realidad aumentada son la combinación de un ambiente virtual con un entorno real. Varios juegos, mediante pantallas, cascos y visores, generan experiencias que mezclan la realidad percibida con elementos virtuales.

Estos juegos están muy lejos del singular soplo divino que habrá de volver conscientes a los robots. Pero hay futurólogos que han imaginado la vida social después del soplo que habría provocado la multiplicación de robots conscientes. Es significativa la proyección que hace un economista de un futuro en el que toda la fuerza de trabajo ha sido sustituida por robots. Robin Hanson en su libro *The Age of Em* explica que dentro de unos cien años, en el siglo XXII, todos los trabajadores humanos habrán sido remplazados por una primera versión de robots construidos a partir de emulaciones de cerebro completo de personas vivas. A partir de este supuesto, intenta dibujar el mundo social y económico resultante.⁸⁷ Es un mundo en el que los humanos ya no trabajan y donde unos robots llamados em (emulaciones) sostienen la economía. Un em es el resultado de copiar de un cerebro humano particular todas sus neuronas y las conexiones que las unen para a continuación construir un sistema computacional capaz de procesar señales de acuerdo con las características copiadas. El resultado es un em que tiene una conducta muy parecida al original, tan parecida que los

humanos pueden hablar con el artificio y convencerlo (u obligarlo) de hacer trabajos útiles. Estos ems habitan en sus ciudades, alejados de los lugares donde los humanos viven una existencia confortable, gozando de las inversiones que han hecho en la economía de los robots. Los robots, por su lado, se organizan en clanes que agrupan a los descendientes del mismo cerebro original. Los ems son entes sexuados que pueden ser felices, estar tristes, sufrir cansancio, tener esperanzas o miedos, hacer amigos y tener amantes. La inmensa mayoría de los ems, más del 80 %, vive una existencia digital y virtual en los circuitos de computadoras, y solamente quienes requieren hacer trabajo físico tienen cuerpos. Pero se trata de cuerpos que no tienen ninguna semejanza con los humanos. Hanson supone que estos androides serán diminutos (256 veces más pequeños que un humano) pero mucho más rápidos tanto física como mentalmente. Es sintomático que en esta versión del futuro los robots aparecen con emociones y sentimientos. De hecho, podrían ser sistemas eficientes carentes de sensibilidad. Pero Hanson ha querido poner un poco de emoción en su versión tecnocrática y aburrida de esa clase de mundos digitales imaginados por Philip K. Dick en sus inquietantes novelas. En vano buscará el lector del libro de Hanson alguna explicación de cómo los humanos han logrado fabricar a unos robots inteligentes y sensibles que trabajan para ellos.

11. LA PRUEBA DEL PLACEBO

Estaremos frente a un robot verdaderamente consciente en el momento en que comprobemos que siente un alivio al aplicarle un placebo cuando sufra un malestar. Ello será la prueba de que está dotado de un sistema al que se puede engañar y que, como resultado, la máquina deja de sentirse enferma. Aunque he expresado muchas dudas sobre la forma en que se intenta construir máquinas inteligentes y dotadas de conciencia, estoy convencido de que sí será posible tener éxito en esta empresa, aunque no me parece que se logre tan pronto como algunos esperan. No se logrará crear robots con conciencia hasta que se resuelva un problema muy complejo: la manera de ensamblar en un solo sistema una inteligencia artificial no especializada de amplio espectro con alguna forma de vida no orgánica capaz de autoorganizarse, autorreplicarse y tener sensibilidad. Dentro de este sistema tendrán que funcionar los equivalentes cibernéticos y mecánicos de un cerebro basado en señales junto con un exocerebro apoyado en símbolos culturales.

La construcción de una inteligencia de carácter general se enfrenta a problemas técnicos muy complicados. Habrá que comenzar sumando varios sistemas especializados que logren coordinarse con agilidad. Ello requerirá de una potencia y una capacidad computacional enormes que no existen actualmente pero que la evolución de la tecnología alcanzará en algún momento. Acaso el perfeccionamiento de las computadoras cuánticas cambie este panorama. La construcción de un robot con una inteligencia flexible, amplia y general capaz de adaptarse y operar en ambientes no estructurados desconocidos se puede llevar a la práctica solamente mediante una máquina capaz de aprender. Un robot realmente inteligente debe aprender mediante la observación de lo que hacen los humanos (u otros robots) o mediante ensayos de prueba y error. El llamado aprendizaje profundo, como ya lo he señalado, está avanzando en esta dirección, pero se encuentra aún muy lejos de alcanzar una capacidad general de aprendizaje multifacético.

La elaboración de formas de vida no orgánica es un problema mucho más

enredado y difícil. Creo que para avanzar en este terreno será necesario abandonar la muy arraigada idea de que la conciencia es independiente del sustrato material que la procesa. Las formas de conciencia que conocemos tienen un sustrato biológico que proporciona la sensibilidad sin la cual no es posible que existan a un nivel superior. De hecho, sin un sustrato biológico sensible tampoco puede haber formas de conciencia primarias como las de un ratón, un gato o un delfín. Estoy pensando en la posibilidad de crear formas no orgánicas de vida, pero durante una etapa de transición se podría experimentar conectando robots con materia orgánica sintética, la que aportaría a la máquina la sensibilidad necesaria para sustentar procesos conscientes semejantes a los humanos. Me gusta suponer que se podrán hacer cultivos de órganos biológicos artificiales capaces de transmitir dolor, placer y otras sensaciones a una poderosa máquina dotada de una inteligencia general. Esta alternativa no es tan tecnológicamente lejana, pues ya se está experimentando con cierto éxito en la creación de órganos diseñados artificialmente, cultivados en animales o en laboratorio. Más complicado será el diseño de un interfaz que los conecte a una máquina y que pueda transmitir sensaciones.

Ahora bien, si pensamos en robots superinteligentes dotados de conciencia que no dependan de las condiciones que conocemos en la Tierra y que puedan sobrevivir a una condición postsolar como la que imaginó Lyotard, en este caso será necesario construir formas de vida artificial no orgánica capaces de organizarse en forma autónoma y de tener sensibilidad. Ya hay algunas reflexiones y pruebas experimentales que muy embrionariamente apuntan hacia esta alternativa. Manuel de Landa ha estudiado fenómenos no orgánicos de autoorganización que ocurren cuando aparece lo que los físicos llaman bifurcaciones.⁸⁸ También se ha intentado crear una especie de vida metálica con polioxometalatos. Pero estamos lejísimos de ver la construcción de formas complejas de vida metálica. Todos los organismos conocidos utilizan compuestos de carbono como base de sus funciones metabólicas y estructurales. Formas alternativas de vida basadas en metaloides como el silicón son una mera especulación.

Los problemas que enfrenta la creación de una conciencia artificial giran en torno de algo que he asumido de manera implícita pero que ahora quiero abordar directamente. Desde mi perspectiva, una máquina consciente debe constituir una individualidad. La conciencia implica la definición de un yo

sensible que se percibe como diferente de otras conciencias con las que convive y relativamente separado del medio ambiente que lo rodea. Pero hay quienes han pensado que en la construcción de máquinas inteligentes existe la posibilidad de robots dotados de una inteligencia colectiva, similar a la que se supone que existe en una colmena. Se ha puesto el ejemplo de Internet como una supuesta conciencia colectiva que funciona como un enjambre. Se asume que en las colonias de insectos hay una especie de alma comunal y que cada insecto carece de individualidad. No sé si existe ese espíritu colectivo en sociedades animales, pero estoy convencido de que, si se construye una red de robots que funcione colectivamente, en ella no habrá una conciencia superior a menos que logre definir límites que la separen de otras redes. Y en este caso habrán surgido las condiciones para una conciencia individual, un ego robótico, una autoconciencia. Esta idea se enfrenta con quienes suponen que la conciencia es una ilusión que no conforma una unidad. Por ejemplo, Francisco Varela, un conocido estudioso de los fenómenos cognitivos, ha expresado que no existe un yo centralizado y unitario; creyó que el budismo está en lo correcto cuando afirma que el yo es una quimera vacía de toda sustancialidad. Para Varela el yo carece de identidad y no es más que un aparente patrón global que emerge de la actividad de componentes simples que pareciera estar localizada centralmente pero que no podemos encontrar en ninguna parte.⁸⁹ El filósofo John Gray ha acogido con gran entusiasmo estas ideas y supone que los humanos no están hechos de una sola pieza: son una sucesión de fragmentos. El yo es una ficción, una ilusión.⁹⁰ Esta idea se encuentra muy arraigada y ha perturbado e incluso frenado la investigación de las bases biológicas y culturales de la conciencia. Aplicada a la inteligencia artificial, esta interpretación considera que un robot no es más que una colección de comportamientos que compiten, que carecen de un sistema central y que solo tienen unidad ante la mirada externa de un observador. Esta es la conclusión de Rodney A. Brooks, un conocido especialista en robótica y computación.⁹¹ Estas ideas descienden de la visión de uno de los fundadores de la inteligencia artificial, Marvin Minsky, quien popularizó la teoría de la «sociedad de mentes», según la cual la mente de los humanos está compuesta de varios agentes autónomos organizados jerárquicamente, desprovistos de inteligencia, especializados e interconectados.

Yo me inclino más bien a señalar con énfasis, como el neurólogo Antonio

Damasio lo ha expresado en todos sus libros, que la homeostasis es un hecho biológico fundamental para entender la conciencia. La homeostasis implica, tanto al nivel elemental de una bacteria como en el funcionamiento del cuerpo humano, la capacidad de regular el medio interior para mantenerlo estable, equilibrado y separado del entorno. Un robot consciente deberá tener la sensibilidad necesaria para mantener homogéneo y estable su medio interno, lo que necesariamente implica la constitución de un sistema individual. La conciencia requiere de una sensibilidad capaz de distinguir el medio interior del ambiente exterior. Los robots suelen ser definidos como «agentes autónomos» capaces de detectar su entorno y actuar de acuerdo con su propia agenda. Son por lo tanto sistemas individuales dotados de una identidad.

Pero el tema de la identidad individual nos enfrenta a una paradoja singular. La conciencia humana genera una identidad individual porque se extiende fuera de los límites del cuerpo gracias a un exocerebro compuesto por prótesis artificiales de carácter simbólico. No se pueden aplicar los mecanismos de la homeostasis de manera rígida, pues aquí nos encontramos con una singularidad: para definir al yo es necesario que la conciencia se expanda hacia los otros que están en nuestro entorno. La conciencia humana necesita mantener un medio interno que al mismo tiempo es externo, pues el cerebro capta la intrusión de elementos exógenos de carácter artificial. Por ello es útil entender que hay otro proceso que complementa la homeostasis y que permite mantener cierto equilibrio frente a la heterogeneidad que aparece en las redes de la conciencia debido a su carácter híbrido. Mientras la homeostasis mantiene el ambiente interior equilibrado en torno de un punto «normal», la heterostasis cambia ese punto de equilibrio cuando el organismo se enfrenta, por ejemplo, a la presencia de sustancias exógenas tóxicas. Esto desencadena un proceso de tolerancia de los agentes externos, con lo cual se establece un nuevo equilibrio. Ello ocurre cuando el organismo se tiene que adaptar a altos niveles de contaminantes venenosos, drogas o alcohol. Se incrementa la tolerancia ante las toxinas y aparece una dependencia. Como lo expliqué en mi libro *Antropología del cerebro*, las prótesis simbólicas e instrumentales se han desarrollado gracias a una dependencia del cerebro, que, además de volverse adicto a ellas, se adaptó también a nuevas situaciones gracias a que establece un nuevo equilibrio con la parte artificial de su entorno. Así pues,

es la heterostasis la que provoca en los humanos lo que he llamado una sociodependencia de los circuitos cerebrales provocada por su incompletitud.⁹²

Estoy convencido de que una conciencia robótica se enfrentará a este mismo problema. Para consolidar su identidad consciente tendrá que abrir y extender sus circuitos internos de tal manera que se genere una interacción entre el sistema cibernético y el entorno cultural de la máquina. Si los robots logran esta hazaña, podrían aparecer individuos androides con inclinaciones difíciles de prever. El experimento acaso engendraría unos extraños monstruos que sería necesario aniquilar o bien podrían surgir seres creativos que serían una compañía benéfica para los humanos del futuro. La ciencia ficción se ha alimentado de estas posibilidades y ha generado mitos muy atractivos. Si finalmente se tiene éxito en la construcción de robots inteligentes y conscientes, seguramente tendrán cierta autonomía y su evolución podría llegar a ser una amenaza. Estos androides serían acaso capaces de replicarse o reproducirse. No sería extraño que surgiese una gama variada de robots conscientes, fruto de diferentes tecnologías asociadas a los grupos o empresas que los habrían creado. Pero estaríamos frente a seres artificiales de diferentes clases y con capacidades distintas. Tendríamos máquinas marcadas por su origen, caracterizadas por una desigualdad primigenia peligrosa. Digamos que el pecado original de estos robots sería el diseño con que habrían sido construidos y que los marcaría para siempre. Las diferentes marcas de robots estarían rodeadas de subculturas exocerebrales diferentes y formarían algo así como grupos tribales dotados de enjambres simbólicos peculiares y distintivos.

Mi imaginación está volando demasiado lejos, pero con este vuelo a un futuro ignoto he querido destacar la importancia de los circuitos exocerebrales en la construcción de robots inteligentes. Sin que tengan entornos simbólicos, sociales y culturales, no podrán construirse máquinas conscientes. Cabe la posibilidad de que los ingenieros creadores de los robots del futuro los conecten a los mismos circuitos exocerebrales que usan los humanos, lo que aminoraría las disparidades impresas en cada marca o especie de máquinas conscientes. Este tema nos lleva directamente al problema de la propiedad de los robots, pues a fin de cuentas serían artefactos construidos por empresas, laboratorios y fábricas. Pero, al ser conscientes y por lo tanto dotados de voluntad, estos aparatos mecánicos no

podrían ser tratados simplemente como los esclavos y los servidores de quienes los han construido o comprado. Ante este problema, hay quienes creen que sería más prudente abandonar los intentos de crear conciencias artificiales. No estoy convencido de que sea lo más sensato, y estoy seguro de que no se detendrá la investigación encaminada a intentar construir conciencias artificiales.

Los impulsos que llevan a crear robots conscientes buscan otro logro similar. Tratan de cambiar el soporte biológico de las conciencias humanas para instalarlas en la maquinaria de un robot. Se trata de un camino paralelo a la construcción de máquinas autónomas con su propia conciencia robótica, pero en este caso se intentaría crear unos artefactos cibernéticos en los que se alojarían las mentes que abandonarían su cuerpo biológico para migrar a una existencia mecánica mucho más prolongada. Podemos suponer que este cambio radical de sustrato implicaría también una profunda transformación de la conciencia tal como la conocemos hoy, y sospecho que estos seres acabarían asemejándose a los androides completamente artificiales. Aquí también podrían producirse desigualdades importantes provocadas por las diferencias en las clases de soporte mecánico que cada persona hubiese logrado adquirir.

Para que los robots alcancen formas de conciencia tan sofisticadas como las humanas, y no sean unos zombis insensibles, deberán pasar por los rituales del placer y el dolor. Sin el sufrimiento de malestares y sin el placer de su alivio es difícil concebir máquinas inteligentes humanoides orientadas por una conciencia desarrollada y compleja. Las mentiras sanadoras del efecto placebo forman parte de una extensa y ramificada cultura que alienta las ilusiones, los mitos y los goces; una cultura que incluye amenazas, frustraciones, engaños y que no es ajena a la magia del poder. Los robots serán conscientes si un chamán del futuro, con el arte de la palabra, logra cautivarlos y se comprueba que son sensibles al efecto placebo.

NOTAS

1 *Antropología del cerebro. Conciencia, cultura y libre albedrío*, Pre-Textos, Valencia, 2014. La primera edición es de 2006.

2 «Más allá del principio del placer» [1920], en *Obras completas*, tomo III, Biblioteca Nueva, Madrid, 1973, p. 2509.

3 Los melquitas fueron los cristianos de Siria, Egipto y Palestina fieles al emperador (Malka, en siríaco) de Bizancio. Originalmente las iglesias melquitas tenían su propio rito, pero acabaron adoptando el ritual bizantino.

4 Giuseppe Gabrieli, «Nota bibliografica su Qust.a- ibn Luqa-», *Rendiconti della Reale Accademia dei Lincei*, Classe di Scienze Morali, Storiche e Filologiche, serie 5, XXI, Roma, 1912-1913, pp. 361ss.

5 Uso la transcripción en latín y la traducción al inglés de Judith Wilcox y John M. Riddle en su estudio «Qust.a- ibn Lu-qa-'s *Physical ligatures* and the recognition of the placebo effect», en *Medieval Encounters*, vol. I, n.º 1: 1-50, junio de 1995.

6 Seguramente el traductor al latín sustituyó Escitia por Escocia. Los textos hipocráticos hacen referencia a los escitas.

7 El término *epode*, en español «epoda» o «epodo», era usado por Homero para referirse a las palabras que tienen un efecto terapéutico. En el *Cármides* se trata de un encantamiento o conjuro. Epodo es el último verso de una estancia repetido muchas veces, y en la poesía griega era la tercera parte de un canto lírico (estrofa, antistrofa y epodo).

8 Véase la reflexión sobre el chamanismo de Manuel Durán «On shamans, magic, and the birth of poetry», *Antemnae* (2001): 39-58.

9 V. M. Mikhailovskii (1846-1904), «Shamanism in Siberia and European Russia», *Journal of the Anthropological Institute of Great Britain and Ireland* 24 (1895): 62-100. Es la traducción del original publicado en 1892 en el volumen 12 de las actas de la Sección de Etnografía de la Sociedad Imperial de Historia Natural, Antropología y Etnografía de Rusia. Sergei Mikhailovich Shirokogoroff (1887-1939), *Psychomental Complex of the Tungus*, Kegan Paul, Londres, 1935.

10 Jean-Paul Roux, «Turkish and Mongolian shamanism», *Mythologies*, compilado por Yves Bonnefoy, University of Chicago Press, Chicago, 1991, vol. 2: 1103-1104.

11 «L'efficacité symbolique», *Revue de l'histoire des religions*, 136, n.º 1 (1949): 5-27.

12 Ernesto de Martino, *Il mondo magico. Prolegomeni a una storia del magismo* [1948], Bollati Boringhieri, Turín, 1973.

13 *Ibid.*, p. 163.

14 «Intorno al "magismo" como età storica», publicado originalmente en *Filosofia e storiografia* (1949), se encuentra incluido en el libro de Ernesto de Martino, pp. 279ss.

15 Mircea Eliade, «Science, idéalisme et phénomènes paranormaux», *Critique* 21 (1948): 315-323; este texto también se encuentra en el libro de Ernesto de Martino. Véase su libro clásico *El chamanismo y las técnicas arcaicas del éxtasis*, Fondo de Cultura Económica, México, 1960.

16 Wilfred Barbrooke Grubb, *An Unknown People in an Unknown Land. An Account of the Life and Customs of the Lengua Indians of the Paraguayan Chaco, with Adventures and Experiences during Twenty Years' Pioneering and Exploration amongst Them*, Seeley, Londres, 1911, pp. 129ss.

17 *La mentalité primitive*, Félix Alcan, París, 1922, capítulo III, p. 2.

- 18 *Les carnets de Lucien Lévy-Bruhl*, Presses Universitaires de France, París, 1949, p. 131.
- 19 David Gordon Wilson, *Redefining Shamanisms: Spiritualist Mediums and Other Traditional Shamans as Apprenticeship Outcomes*, Bloomsbury, Londres, 2013. También puede consultarse el punto de vista de una chamán estadounidense, no indígena, en la enciclopedia preparada por Christina Pratt, *An Encyclopedia of Shamanism*, Rosen, Nueva York, 2007. La autora encabeza el Last Mask Center for Shamanic Healing en Nueva York.
- 20 No uso la traducción de 1947 en la que se basó LéviStrauss, sino la versión completa de 1953: Nils M. Holmer y S. Henry Wassén, *The Complete Mu-Igala in Picture Writing. A Native Record of a Cuna Indian Medicine Song*, Etnologiska Studier, 21, Goteborg, 1953.
- 21 Aparentemente el niño está a punto de salir o ya ha salido, pero la placenta no es expulsada. Esta es la interpretación de Staffan Mjönes, «Shaman, psychoanalyst or obstetrician: A critical reading of Claude Lévi-Strauss' essay "The efficiency of symbols"», *Folklore* 45 (2010): pp. 7-26.
- 22 Lévi-Strauss conoció solamente hasta este punto el canto, pues todavía no se publicaba la última parte en la edición que él consultó. Para la reedición de su ensayo no se tomó la molestia de consultar la versión completa de 1953. Creyó equivocadamente que ya se había logrado el parto, aunque no se describía en el texto kuna que él conoció. Tampoco conoció la versión pictográfica. Además, según explica Staffan Mjönes, faltaba lograr que la mujer expeliera la placenta, cuya retención era probablemente la razón de su malestar (Mjönes, *op. cit.*).
- 23 Michel Perrin, «Las artes de los chamanes», en *Chamanismo: Tiempo y lugares sagrados* (Fernando Barona Tovar, compilador), Universidad del Valle, Cali, 2007.
- 24 Michael Taussig, *Shamanism, Colonialism, and the Wild Man. A Study in Terror and Healing*, University of Chicago Press, Chicago, 1987, pp. 389s.
- 25 M. A. Secheyay, *La realización simbólica. Diario de una esquizofrénica*, Fondo de Cultura Económica, México, 1958.
- 26 Fabrizio Benedetti, Elisa Carlino y Antonella Pollo, «How placebos change the patient's brain», *Neuropsychopharmacology* 36 (2011): 339-354.
- 27 Véase Henry K. Beecher, «The powerful placebo», *Journal of the American Medical Association* 159 (1955): 1602-1606.
- 28 Anne Harrington, *The Cure Within: A History of Mindbody Medicine*, Norton, Nueva York, 2008, pp. 126-127.
- 29 Fabrizio Benedetti, Elisa Carlino y Antonella Pollo, *op. cit.*, p. 348.
- 30 *Ibid.*, pp. 349-350.
- 31 Oron Frenkel, «A phenomenology of the "placebo effect". Taking meaning from the mind to the body», *Journal of Medicine and Philosophy* 33 (2008): 58-79.
- 32 Jennifer Jo Thompson, Cheryl Ritenbaugh y Mark Nichter, «Reconsidering the placebo response from a broad anthropological perspective», *Culture, Medicine and Psychiatry* 33 (2009): 112-152.
- 33 Daniel E. Moerman, *Meaning, Medicine, and the «Placebo» Effect*, Cambridge University Press, Cambridge, 2002, p. 13.
- 34 Franklin G. Miller, Luana Colloca y Ted J. Kaptchuk, «The placebo effect: illness and interpersonal healing», *Perspectives in Biology and Medicine* 52, 4 (2009): 1-20.
- 35 Fabrizio Benedetti, «Placebo and the new physiology of the doctor-patient relationship», *Physiological Reviews* 93 (2013): 1207-1246.
- 36 Irving Kirsch, «Antidepressants and the placebo effect», *Zeitschrift für Psychologie* 222, 3 (2014): 128-134.
- 37 Pim Cuijpers y Ioana A. Cristea, «What if placebo effect explained all the activity of depression treatments?», *World Psychiatry* 14, 3 (2015): 310-311.
- 38 Fabrizio Benedetti y Aziz Shaibani, «Nocebo effects: more investigation is needed», *Expert*

Opinion on Drug Safety 17, n.º 6 (mayo de 2018): 541-543.

39 J. D. Levine, N. C. Gordon y H. L. Fields, «The mechanism of the placebo effect», *The Lancet* 312, n.º 8091 (23 de septiembre de 1978): 654-657. Fabrizio Benedetti, H. S. Mayberg, T. D. Wagner, C. S. Stohler y J. K. Zubieta, «Neurobiological mechanisms of the placebo effect», *The Journal of Neuroscience* 25 (2005): 10390-10402.

40. J. J. Lipman, B. E. Miller, K. S. Mays, M. N. Miller, W. C. North, W. L. Byrne y B. Peak, «Endorphin concentration in cerebrospinal fluid: reduced in chronic pain patients and increased during the placebo response», *Psychopharmacology* 102 (1990): 112-116.

41 Alina Dimitriu y Bogdan O. Popescu, «Placebo effects in neurological diseases», *Journal of Medicine and Life* 3, n.º 2 (2010): 114-121.

42 Fabrizio Benedetti, «Drugs and placebos: what's the difference?», *EMBO Reports* 15, n.º 4 (2014): 329-332.

43 Rasha Al-Lamee *et al.*, «Percutaneous coronary intervention in stable angina (ORBITA): a double-blind, randomized controlled trial», *The Lancet* 391, n.º 10115 (6 de enero de 2018): 31-40.

44 Véase una buena descripción general con muchos ejemplos en los libros ya citados de Daniel E. Moerman, *Meaning, Medicine, and the «Placebo» Effect*, y de Anne Harrington, *The Cure Within. A History of Mind-body Medicine*. Véase igualmente una excelente perspectiva general en el libro de Alain Autret *Les effets placebo. Des relations entre croyances et médecines*, L'Harmattan, París, 2013.

45 Antonio Damasio, *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, Harcourt Brace, Nueva York, 1999, pp. 51 y 60. Véase también del mismo autor *The Strange Order of Things. Life, Feeling, and the Making of Culture*, Pantheon Books, Nueva York, 2018.

46. Sobre el efecto nocebo véase Alexandra Tinnermann, S. Geuter, C. Sprenger, J. Finsterbusch y C. Büchel, «Interactions between brain and spinal cord mediate value effects in nocebo hyperalgesia», *Science* 6359 (6 de octubre de 2017): 105-108.

47 Russell W. Belk, «Possessions and the extended self», *Journal of Consumer Research* 15 (1988): 139-168.

48 Russell B. Clayton, Glenn Leshner y Anthony Almond, «The extended iSelf: The impact of iPhone separation on cognition, emotion, and physiology», *Journal of Mediated Communication* 20 (2015): 119-135.

49 Dan-Mikael Ellingsen, Johan Wessberg, Marie Eikemo, Jaquette Liljencrantz, Tor Endestad, Håkan Olausson y Siri Leknos, «Placebo improves pleasure and pain through opposite modulation on sensory processing», *PNAS* 110 (2013): 1799317998.

50 Sigmund Freud, «Psicoterapia (tratamiento por el espíritu)», *Obras completas*, Biblioteca Nueva, Madrid, 1973, tomo I, p. 1020.

51 Sigmund Freud, «Sobre psicoterapia», *op. cit.*, p. 1008.

52 *Ibid.*, p. 1013.

53 Stevan Harnad, «Why and how we are not zombies», *Journal of Consciousness Studies* 1 (1995): pp. 164-167.

54 Agustín de Hipona, *La ciudad de Dios*, 14:23:3.

55 Gerald Edelman, *Wider Than the Sky. The Phenomenal Gift of Consciousness*, Yale University Press, New Haven, 2004, pp. 80, 145 y 180.

56 Ray Kurzweil, *The Singularity Is Near: When Humans Transcend Biology*, Viking, Nueva York, 2005. Véase la muy divertida e irónica exploración del transhumanismo escrita por Mark O'Connell: *To Be a Machine. Adventures Among Cyborgs, Utopians, Hackers, and the Futurists Solving the Modest Problem of Death*, Granta, Londres, 2017.

57 Susan Blackmore, «Consciousness in Meme Machines», en *Machine Consciousness*, editado

por Owen Holland, *Journal of Consciousness Studies* 10 (4-5), 2003.

58 Susan Blackmore, *Conversations on Consciousness: What the Best Minds Think About the Brain, Free Will, and What It Means To Be Human*, Oxford University Press, Oxford, 2006, p. 203.

59 Peter Farleigh, «The Ensemble and the Single Mind. A Sceptical Inquiry», en *Artificial Consciousness*, Antonio Chella y Riccardo Manzotti, eds., Imprint Academic, Exeter, 2007.

60 Véase Roger Penrose, *The Emperor's New Mind*, Oxford University Press, Oxford, 1989.

61 Véase Alan Winfield, *Robotics*, Oxford University Press, Oxford, 2012, p. 8. Una buena introducción general a la robótica ha sido publicada por Laurence Devillers, *Des robots et des hommes. Mythes, fantômes et réalité*, Plon, París, 2017.

62 *Ibid.*, pp. 51-54.

63 Véase Stan Franklin, «IDA: A Conscious Artifact?», en *Machine Consciousness*, editado por Owen Holland, *Journal of Consciousness Studies* 10 (4-5), 2003.

64 *Ibid.*, p. 64.

65 Alan Winfield, *Robotics*, *op. cit.*, p. 132.

66. Vincenzo Tagliascio, «Artificial Consciousness: A Technological Discipline», en *Artificial Consciousness*, Antonio Chella y Riccardo Manzotti, eds., *op. cit.*

67 Murray Shanahan, *The Technological Singularity*, MIT Press, Cambridge, Mass., 2015.

68 El lector puede ver cómo funciona DeepMind en [https:// tinyurl.com/atariai](https://tinyurl.com/atariai).

69 Apurv Mishra, «AI that sees like humans», *Scientific American*, diciembre de 2017.

70 Murray Shanahan, *The Technological Singularity*, MIT Press, Cambridge, Mass., 2015, p. 3.

71 Terrence J. Sejnowski, *The Deep Learning Revolution*, MIT Press, Cambridge, Mass., 2018.

72 Véase una buena panorámica de la inteligencia artificial en el libro de Margaret A. Boden *AI: Its Nature and Future*, Oxford University Press, Oxford, 2016.

73 Max Tegmark, *Life 3.0. Being Human in the Age of Artificial Intelligence*, Knopf, Nueva York, 2017, pp. 300-305. Puede leerse una crítica a las tesis de Giulio Tononi en mi libro *Antropología del cerebro*, pp. 273-276.

74 Christof Koch, *Consciousness: Confessions of a Romantic Reductionist*, MIT Press, Cambridge, Mass., 2012, p. 118.

75 *Ibid.*, p. 311.

76 Michio Kaku, *The Future of the Mind*, Anchor Books, Nueva York, 2015, p. 334.

77 Murray Shanahan, *The Technological Singularity*, *op. cit.*, pp. 6-10.

78 Véase Jerry Fodor, *The Language of Thought*, Harvard University Press, Cambridge, Mass., 1975. Una panorámica general de la idea de que la mente es una máquina y de la teoría computacional de la cognición puede consultarse en Tim Crane, *The Mechanical Mind. A Philosophical Introduction to Minds, Machines and Mental Representation*, 3.^a edición, Routledge, Londres, 2016.

79 La ironía de Marx aparece en el último capítulo del primer tomo de *El capital*, dedicado a las teorías modernas de la colonización, y se basa en lo que cuenta el gran exponente del «arte de la colonización», Edward Gibbon Wakefield.

80 Murray Shanahan, *The Technological Singularity*, *op. cit.*, pp. 122ss.

81. Michael S. A. Graziano, «The Attention Schema Theory. A Foundation for Engineering Artificial Consciousness», *Frontiers in Robotics and AI* 4, noviembre de 2017. Véase también su libro *Consciousness and the Social Brain*, Oxford University Press, Oxford, 2013.

82 Jun Tani, *Exploring Robotic Minds. Actions, Symbols, and Consciousness as Self-Organizing Dynamic Phenomena*, Oxford University Press, Oxford, 2017.

83 Véase Thomas Heimburg, «Phase transitions in biological membranes», *arXiv.org*: 1805.11481v1, 2018. Thomas Heimburg y Andrew J. Jackson, «On soliton propagation in biomembranes and nerves», *PNAS* 102, núm. 28 (2005): 97909795. Y el estimulante comentario

sobre la comunicación solitónica de Douglas Fox, «The brain, reimagined», *Scientific American* 318, núm. 4 (2018): 53-59. Otros estudios plantean tesis similares, como Ahmed El Hady y Benjamin B. Machta, «Mechanical surface waves accompany action potential propagation», *Nature Communications* 6:6697 doi: 10.1038/ncomms 7697 (2015).

84 Ray Kurzweil, *The Singularity Is Near*, *op. cit.*

85 Jean-François Lyotard, «Si l'on peut penser sans corps», en *L'inhumain. Causeries sur le temps*, Galilée, París, 1988. Este texto recoge una conferencia impartida en un seminario de la universidad alemana de Siegen, donde fue invitado por Hans Ulrich Gumbrecht.

86 Los astrónomos hoy piensan que el sistema solar durará todavía seis o siete mil millones de años, o incluso más.

87 Robin Hanson, *The Age of Em. Work, Love, and Life when Robots Rule the Earth*, Oxford University Press, Oxford, 2016.

88 Manuel de Landa, «Non Organic Life», *Incorporations*, Jonathan Crary y Sanford Kwinter, eds., Zone, Nueva York, 1992.

89 Francisco J. Varela, *Ethical Know-How: Action, Wisdom, and Cognition*, Stanford University Press, Stanford, 1999.

90 John Gray, *Straw Dogs: Thoughts on Humans and Other Animals*, Farrar, Strauss and Giroux, Nueva York, 2002.

91 Rodney A. Brooks, «Achieving Artificial Intelligence through Building Robots», *A.I. Memo* 899, Cambridge Artificial Intelligence Laboratory, 1986.

92 *Antropología del cerebro*, *op. cit.*, pp. 335ss.

Edición en formato digital: septiembre de 2019

© imagen de cubierta, Armando Veve, «Your Conscious Unconscious», 2016

© Roger Bartra, 2019

© EDITORIAL ANAGRAMA, S.A., 2019

Pedró de la Creu, 58
08034 Barcelona

ISBN: 978-84-339-4091-9

Conversión a formato digital: Newcomlab, S.L.

anagrama@anagrama-ed.es

www.anagrama-ed.es